

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2004, 07:39:00 ; Search time 7452 Seconds

(without alignments)  
16919.604 Million cell updates/sec

Title: US-09-700-712a-1

Perfect score: 2909

Sequence: 1 gagaagatttcgtatcgcctc.....cagcgaatcggcgctgcag 2909

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

```
Database :
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_ph:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_stb:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pac:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_stb:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_hum:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pin:*
35: em_htg_rtd:*
36: em_htg_nam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2909	100.0	2909	1	VCH6514	AJ006514 Vibrio ch
2	2909	100.0	2909	6	AX009481	AX009481 Sequence
3	2870.6	98.7	9931	1	AE004153	AE004153 Vibrio ch
4	1390.8	47.8	1410	1	VCH714G	Y17135 Vibrio chol
5	1222	42.0	1222	6	AX009483	AX009483 Sequence
6	1216.8	41.8	1236	1	AY143429	AY143429 Vibrio ch
7	1187	40.8	1591	1	VCA110968	AJ010968 Vibrio ch
8	1052.4	36.2	298900	1	AP005074	AP005074 Vibrio pa
9	1035	35.6	301442	1	AE016798	AE016798 Vibrio pa
10	1028.6	35.4	249150	1	AP005332	AP005332 Vibrio vu
11	841.4	28.9	852	6	AX020429	AX020429 Sequence
12	838	28.8	838	6	AX009482	AX009482 Sequence
13	814.8	28.0	10719	1	AE006043	AE006043 Pasteurel
14	775.2	26.6	11545	1	AP064791	AP064791 Pasteurel
15	744.2	25.6	11545	1	U32772	U32772 Haemophilus
16	744.2	25.6	110000	6	AR274513_09	AR274513 Haemophil
17	433.8	14.1	873	6	AE017153	AE017153 Haemophil
18	410.8	14.1	873	6	AX294152	AX294152 Pirellula
19	352.2	12.1	304250	1	AE015577	AE015577 Shewanell
20	329.4	11.3	10687	1	AE017151	AE017151 Haemophil
21	314.8	10.8	304558	1	AE017151	AE017151 Haemophil
22	292.2	10.0	807	6	AR139598	AR139598 Sequence
23	280.6	9.7	10474	1	AE004472	AE004472 Pseudomon
24	280.6	9.6	10029	1	AE013918	AE013918 Yersinia
25	280.6	9.6	208050	1	AJ414144	AJ414144 Yersinia
26	278.2	9.6	342905	1	BX571861	BX571861 Phototrab
27	278.2	9.6	349980	6	AX770908	AX770908 Sequence
28	273	9.4	10242	1	AE005511	AE005511 Escherich
29	273	9.4	266558	1	AP002563	AP002563 Escherich
30	265.8	9.1	891	6	AR377359	AR377359 Sequence
31	264.2	9.1	305325	1	AE016765	AE016765 Escherich
32	261.4	9.0	310613	1	AE016875	AE016875 Pseudomon
33	260.8	9.0	10405	1	AE000366	AE000366 Escherich
34	260.2	8.9	12943	1	AE015298	AE015298 Shigella
35	260.2	8.9	292906	1	AE016988	AE016988 Shigella
36	259.6	8.9	71128	1	ECU29581	ECU29581 Escherichia
37	258	8.9	1473	1	STY1GTX	L13259 Salmonella
38	258	8.9	21583	1	AE008838	AE008838 Salmonell
39	256.8	8.8	879	6	AR386236	AR386236 Sequence
40	256.4	8.8	230050	1	AL627277	AL627277 Salmonell
41	256.4	8.8	301311	1	AE016843	AE016843 Salmonell
42	248.2	8.5	876	6	AX188981	AX188981 Sequence
43	248.2	8.5	1310	1	ECOLGT	U12289 Escherichia
44	240.8	8.3	301086	1	AE016964	AE016964 Coxidia
45	232.6	8.0	303438	1	AE017212	AE017212 Geobacter

#### ALIGNMENTS

```
RESULT 1
VCH6514
LOCUS
DEFINITION
Vibrio cholerae lgt and thya genes.
ACCESSION
AJ006514
VERSION
AJ006514.1 GI:3201563
KEYWORDS
lgt gene; prolipoprotein diacylglycerol transferase; thya;
thymidylate synthetase.
SOURCE
Vibrio cholerae
ORGANISM
Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
Vibrionaceae; Vibrio.
REFERENCE
Carlin, N.I.A., Nilsson, A., Todorovic, M., Holmgren, J. and Lebens, M.
AUTHORS
TITLE
Characterisation of the lgt/thya locus from Vibrio cholerae
```

JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2909)  
AUTHORS Carlin,N.I.A.  
TITLE Direct Submission  
JOURNAL Submitted (29-MAY-1998) Carlin N.I.A., Department of Molecular  
Biology, SBL Vaccin AB, PMB, UV 30, Stockholm, S-105 21, SWEDEN  
Location/Qualifiers  
1. 2909  
FEATURES  
/organism="Vibrio cholerae"  
/mol\_type="genomic DNA"  
/strain="US1569"  
/db\_xref="taxon:666"  
14. .829  
gene  
/gene="1gt"  
14. .829  
CDS  
/codon\_start=1  
/transl\_table=11  
/product="prolipoprotein diacylglycerol transferase"  
/protein\_id="CAA07072.1"  
/db\_xref="GI:3201564"  
/db\_xref="GOA:O87077"  
/db\_xref="SPTREMBL:O87077"  
/translation="MPQYLQPNIDPVLFSIGPLAVRWYGLMTLVGELFAMWLANR  
ADRAGSGWTRQVSDILFAGFLGVVIGRWVIVFYNEDLFLADPLYLEKWTGMSF  
HGLGLVITAMFWYARKNQTFFGVADFWAPLPPGLGMRIGNFMNSLMGRVTDVP  
WAFVFNNGEPLPRHPSQLYEFALKEGVLEFLINWFLGKRPGLGSVSGLEFLAGYTFRF  
LVEVYRDPDAQGLFGFGLSMGQILSLPWHVIGILMVMWSYRGLYQDRVAK"  
669. .775  
gene  
/gene="thya"  
669. .775  
-35\_signal  
/gene="thya"  
691. .696  
-10\_signal  
/gene="thya"  
Join(828. .832, 839. .1690, 1725. .1765)  
RBS  
/gene="thya"  
828. .832  
gene  
/gene="thya"  
839. .1690  
CDS  
/codon\_start=1  
/evidence="experimental"  
/transl\_table=11  
/product="thymidylate synthetase"  
/protein\_id="CAA07073.1"  
/db\_xref="GI:3201565"  
/db\_xref="GOA:O86108"  
/db\_xref="SWISS-PROT:O66108"  
/translation="MKQYLDLCORIVDQGVWVENERTGRKCLTVINADLTVDVGNNOF  
PLVTRKSFMAVAELGDIRGYDNAADPRLQKTMNDANLNQAMLNPNYRKGED  
DMGRVYGVCGRAKRPDGHIDOLKKIYVDLSRGVDDREILNPNPGEFHHGCHRPC  
MYSHHSILGDLTVLNSYORSQDPLGALNPNVQYVPLAQAQITGKKRGLAYKIV  
NAHYODQLMRDVOLKREPPAPQFHPINPKIKTIDLETVTDDPVTGYQFHPD  
10YPSV"  
1082. .1084  
variation  
/gene="thya"  
/note="In trimethoprim resistant mutant US1569 4.4, stop  
codon"  
/replace="tag"  
1725. .1765  
stem\_loop  
/gene="thya"  
ORIGIN  
Query Match 100.0%; Score 2909; DB 1; Length 2909;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2909; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GAGAGCTTTGATACCTCAGGCTATCGAGTTTCCCAATATTGACCCGATATTGT 60  
Db 1 GAGAGCTTTGATACCTCAGGCTATCGAGTTTCCCAATATTGACCCGATATTGT 60  
QY 61 TTCGATCGGCCCTCTAGCGGTGCGCTGATAGCTTATTTGGTGGGTTCCCTTT 120  
Db 1 TTCGATCGGCCCTCTAGCGGTGCGCTGATAGCTTATTTGGTGGGTTCCCTTT 120

Db 61 TTCGATCGGCCCTCTAGCGGTGCGCTGATAGCTTATTTGGTGGGTTCCCTTT 120  
QY 121 TCGTATGCTGTGCGCAATCCGCCGAGCGGATCGCGCGAGTGGTGAACGCGTAGCA 180  
Db 121 TCGTATGCTGTGCGCAATCCGCCGAGCGGATCGCGCGAGTGGTGAACGCGTAGCA 180  
QY 181 AGCTCTGACTTGTATTTCGCCGCTTTTATAGTGTATGATTCGGTGGCCGAGTTGGTTA 240  
Db 181 AGCTCTGACTTGTATTTCGCCGCTTTTATAGTGTATGATTCGGTGGCCGAGTTGGTTA 240  
QY 241 TGTGATCTTCAACAATTTGATCTGTTCCTGCTGACCCCTCTTATTTCAAAAGTGG 300  
Db 241 TGTGATCTTCAACAATTTGATCTGTTCCTGCTGACCCCTCTTATTTCAAAAGTGG 300  
QY 301 GACTGCGCGCAATGTCCTTCAACGCGGCTTATTTGGTGTGATCAACGCGCATGTTCTGTA 360  
Db 301 GACTGCGCGCAATGTCCTTCAACGCGGCTTATTTGGTGTGATCAACGCGCATGTTCTGTA 360  
QY 361 TGCAGGTAAAAACCAACGACCTTCTTGTGTGTGCGCATTTTGTGCCCCCTTTAGTCC 420  
Db 361 TGCAGGTAAAAACCAACGACCTTCTTGTGTGTGCGCATTTTGTGCCCCCTTTAGTCC 420  
QY 421 ATTGCGTTTGGGATGAGGACGATTCGGTATCTTATGAAATGATGAATTTGGGGACGAGT 480  
Db 421 ATTGCGTTTGGGATGAGGACGATTCGGTATCTTATGAAATGATGAATTTGGGGACGAGT 480  
QY 481 AACGATATGCTTGGGCTTTTGTATTCCTTAATGTTGGCCCACTGCCGCGCATCTTTC 540  
Db 481 AACGATATGCTTGGGCTTTTGTATTCCTTAATGTTGGCCCACTGCCGCGCATCTTTC 540  
QY 541 ACAGCTTTATGAAATTCGCTTGAAGAGCGGTGCTGTCTTTATCTTAAATGCTTTAT 600  
Db 541 ACAGCTTTATGAAATTCGCTTGAAGAGCGGTGCTGTCTTTATCTTAAATGCTTTAT 600  
QY 601 TGTGTAACCTGTCGCTGAGGAGGATCCGGAATGTTTATGCTGATACGGTACAT 660  
Db 601 TGTGTAACCTGTCGCTGAGGAGGATCCGGAATGTTTATGCTGATACGGTACAT 660  
QY 661 CCGCTTCCCTTGGAAATAGTCCGAGACAGATCTCAGTTGGGCTGTTTGGTGGCT 720  
Db 661 CCGCTTCCCTTGGAAATAGTCCGAGACAGATCTCAGTTGGGCTGTTTGGTGGCT 720  
QY 721 CATTTCAATGGGCAATCTTCTCTTACCTATGATGATCAATCGGATTTTGAATGATG 780  
Db 721 CATTTCAATGGGCAATCTTCTCTTACCTATGATGATCAATCGGATTTTGAATGATG 780  
QY 781 TTGGTCTTACAGACCGCGTTGTATCAGAACCGGTACACGAATAATGGGTAGTAGT 840  
Db 781 TTGGTCTTACAGACCGCGTTGTATCAGAACCGGTACACGAATAATGGGTAGTAGT 840  
QY 841 GAAACAGTATTAGATCTTTGTCAAGCGATCGTCAAGAGTGTGGTGGTGAATAATGA 900  
Db 841 GAAACAGTATTAGATCTTTGTCAAGCGATCGTCAAGAGTGTGGTGGTGAATAATGA 900  
QY 901 ACGAACGGGCAACGCTTGTGTGACTGTGATTAATGCCATTTGACCTTACGATGGGCA 960  
Db 901 ACGAACGGGCAACGCTTGTGTGACTGTGATTAATGCCATTTGACCTTACGATGGGCA 960  
QY 961 CAATCAGTTTCTCTAGTGAATCAACGCAAGTTTTGGAAAGTGGCCGTAACCGGAGT 1020  
Db 961 CAATCAGTTTCTCTAGTGAATCAACGCAAGTTTTGGAAAGTGGCCGTAACCGGAGT 1020  
QY 1021 GCTGCGCTATATTCGTGTGATGATTAAGCGCGGATTTTGGCCATTAAGTACCAAAAC 1080  
Db 1021 GCTGCGCTATATTCGTGTGATGATTAAGCGCGGATTTTGGCCATTAAGTACCAAAAC 1080  
QY 1081 CTGGGATCTATATGCAATTTAAACCAAGCATGGCTCAACATCTTTACCGTAAGAGTGA 1140  
Db 1081 CTGGGATCTATATGCAATTTAAACCAAGCATGGCTCAACATCTTTACCGTAAGAGTGA 1140  
QY 1141 GGATGACATGGGACCGGTGATGTGTTCAGGGTGAAGCTTGGGCTAAGCCTGATGTGG 1200  
Db 1141 GGATGACATGGGACCGGTGATGTGTTCAGGGTGAAGCTTGGGCTAAGCCTGATGTGG 1200

QY	1201	TCATATTGACCAAGCTGGAAAAAGATTGTTGATGATTTGAGCCGTGGCGTGTATGACCGAAG	1260
Db	1201	TCATATTGACCAAGCTGGAAAAAGATTGTTGATGATTTGAGCCGTGGCGTGTATGACCGAAG	1260
QY	1261	TGAAATTCCTAACTCTCACAATCCCGGGTGAATTTCCATGAGGAGTGTTGCCGCTTGAT	1320
Db	1261	TGAAATTCCTAACTCTCACAATCCCGGGTGAATTTCCATGAGGAGTGTTGCCGCTTGAT	1320
QY	1331	GTACAGCCATCATTTTTCATTGCTGGGGGATACCTTGATCTCAACAGTACTCAGCGTTC	1380
Db	1331	GTACAGCCATCATTTTTCATTGCTGGGGGATACCTTGATCTCAACAGTACTCAGCGTTC	1380
QY	1361	ATGTGATGTGCCCTTGGGGTGGAAATTTCAACATGCTGAGAGTTTATGTGTTCCCTGGCT	1440
Db	1361	ATGTGATGTGCCCTTGGGGTGGAAATTTCAACATGCTGAGAGTTTATGTGTTCCCTGGCT	1440
QY	1441	GATGSCACAGATCACAGGGAAAAAGCCGGGCTTGCGCTGATCACAGATCGTCAATGCGCA	1500
Db	1441	GATGSCACAGATCACAGGGAAAAAGCCGGGCTTGCGCTGATCACAGATCGTCAATGCGCA	1500
QY	1501	CATTATACCAAGATCAACTCGAATTGATGCGGATGTGCAAGCTAAACGTGACCATTTCC	1560
Db	1501	CATTATACCAAGATCAACTCGAATTGATGCGGATGTGCAAGCTAAACGTGACCATTTCC	1560
QY	1561	AGCGCTCAGTTCCATATCAATCCAAAGATTAAACACTGACGATTTGGAACTTGAGT	1620
Db	1561	AGCGCTCAGTTCCATATCAATCCAAAGATTAAACACTGACGATTTGGAACTTGAGT	1620
QY	1621	CACCTTGGATGATTTTGAAGTCAACCGGATATCAGTCCAGATCCATTAATACCCGTT	1680
Db	1621	CACCTTGGATGATTTTGAAGTCAACCGGATATCAGTCCAGATCCATTAATACCCGTT	1680
QY	1661	TTCACTCTAATCCCGTATTCAGCGCGGTATGCTTGATGGTATTATATAAAAAAGCTCC	1740
Db	1661	TTCACTCTAATCCCGTATTCAGCGCGGTATGCTTGATGGTATTATATAAAAAAGCTCC	1740
QY	1741	CGAAGCTCGGAGACTTTTATATACAGATGACTTTAAGCTTAAGCGGTTAAGGCGCAAG	1800
Db	1741	CGAAGCTCGGAGACTTTTATATACAGATGACTTTAAGCTTAAGCGGTTAAGGCGCAAG	1800
QY	1801	AATGCTGCCGGGATGAGACAAACACGCCAATTAAGTATCTCACACACACATTTGCT	1860
Db	1801	AATGCTGCCGGGATGAGACAAACACGCCAATTAAGTATCTCACACACACATTTGCT	1860
QY	1861	CTTACAAACCCCAAGTTAGATGAGCTCAGACCTTTAATAGAGATTCGCGCTAAGAAAG	1920
Db	1861	CTTACAAACCCCAAGTTAGATGAGCTCAGACCTTTAATAGAGATTCGCGCTAAGAAAG	1920
QY	1921	AATACCGTAAATCAAGACCGTATGACCATCAAGTTAAAGCTTAAGTGCACACAGCGCAATTTG	1980
Db	1921	AATACCGTAAATCAAGACCGTATGACCATCAAGTTAAAGCTTAAGTGCACACAGCGCAATTTG	1980
QY	1961	CAGAGCAAAACACGGGAAACTCACCCAGAGACACGGGTTCGGCGACGACGATTAATCA	2040
Db	1961	CAGAGCAAAACACGGGAAACTCACCCAGAGACACGGGTTCGGCGACGACGATTAATCA	2040
QY	2041	AGTGCATATGTTCCGACCTTAAGGTAAAGGGTAGATTTCAAGCACTTTACAGCAGCCAGA	2100
Db	2041	AGTGCATATGTTCCGACCTTAAGGTAAAGGGTAGATTTCAAGCACTTTACAGCAGCCAGA	2100
QY	2101	GCCCAACGAGAGAAACCATTAGGCTGTGTGTGTCGATGAAGATTGAACCTAATACCGTAAC	2160
Db	2101	GCCCAACGAGAGAAACCATTAGGCTGTGTGTGTCGATGAAGATTGAACCTAATACCGTAAC	2160
QY	2161	CACGTGACTGAGCAATACCGTGTATGTGGCCCTCGCGCAATCCCATTTTGTAAATTTTC	2220
Db	2161	CACGTGACTGAGCAATACCGTGTATGTGGCCCTCGCGCAATCCCATTTTGTAAATTTTC	2220
QY	2221	ACGTGCGGGGCAACCATCAAACTCTTATAGTTTGCCCATCACCGTAATGGGAGAGAA	2280
Db	2221	ACGTGCGGGGCAACCATCAAACTCTTATAGTTTGCCCATCACCGTAATGGGAGAGAA	2280

QY	2281	AATGGTCGCAATACCACATACGATTAAGGCGACACCAACCGAAAGATTAATACCAATACCGA	2340
Db	2281	AATGGTCGCAATACCACATACGATTAAGGCGACACCAACCGAAAGATTAATACCAATACCGA	2340
QY	2341	AAGCTGGGTTTCAAGCCCTGTGATGACAGGTTTGGTAAATCGGTTTGATTAATAATCAAAACC	2400
Db	2341	AAGCTGGGTTTCAAGCCCTGTGATGACAGGTTTGGTAAATCGGTTTGATTAATAATCAAAACC	2400
QY	2401	TTTCATGCTCATTCGCGACATGCGAAGAGAGAGGGGAAACGAGCCAGCTGTGACCTTCTC	2460
Db	2401	TTTCATGCTCATTCGCGACATGCGAAGAGAGGGGAAACGAGCTGTGACCTTCTC	2460
QY	2461	TAAATGCGCAACATCATTTTCTTAGAGGTAGAGAAATCAGCACCCGCGAGAAATTGAAAAA	2520
Db	2461	TAAATGCGCAACATCATTTTCTTAGAGGTAGAGAAATCAGCACCCGCGAGAAATTGAAAAA	2520
QY	2521	ATCGTGATGTGGCACTGCGGAAAGCACGCGGAAACTCTTTCTTACAGGCGATATGGCC	2580
Db	2521	ATCGTGATGTGGCACTGCGGAAAGCACGCGGAAACTCTTTCTTACAGGCGATATGGCC	2580
QY	2581	AAGGTGACGAGATTTTGGTCAACGTAATCCAAATATTGGCACCCATCACCATATAGAAAT	2640
Db	2581	AAGGTGACGAGATTTTGGTCAACGTAATCCAAATATTGGCACCCATCACCATATAGAAAT	2640
QY	2641	CGCGGTTTCAACCGGTAAACCAACGCGCAACGAGACCAACATTAATAGAAATCAGCGTCT	2700
Db	2641	CGCGGTTTCAACCGGTAAACCAACGCGCAACGAGACCAACATTAATAGAAATCAGCGTCT	2700
QY	2701	TGAGGATTTGAATCAAGTGCAGTGGCACTAAACCAATCATCATCTCGCAATTGGGTGGGA	2760
Db	2701	TGAGGATTTGAATCAAGTGCAGTGGCACTAAACCAATCATCATCTCGCAATTGGGTGGGA	2760
QY	2761	AGCAAAATTCAAATGAACCTTTGGCTTGATGCGCGGTTGCCATTTAAACCGCGCCGAC	2820
Db	2761	AGCAAAATTCAAATGAACCTTTGGCTTGATGCGCGGTTGCCATTTAAACCGCGCCGAC	2820
QY	2821	CATCGCGCACTCAAGAGTAGTAATATACAGCATGAAAGCCAAAGTTTGCCCAACGTAGGCC	2880
Db	2821	CATCGCGCACTCAAGAGTAGTAATATACAGCATGAAAGCCAAAGTTTGCCCAACGTAGGCC	2880
QY	2881	TTTCGTGGTCAAGCAAAATCGCGCTGTGACG	2909
Db	2881	TTTCGTGGTCAAGCAAAATCGCGCTGTGACG	2909
RESULT 2			
AX009481			
LOCUS	AX009481	2909 bp	DNA
DEFINITION	Sequence 1 from Patent WO961634.	linear	PAT 06-SEP-2000
ACCESSION	AX009481		
VERSION	AX009481.1	GI:9996766	
KEYWORDS	Vibrio cholerae		
SOURCE	Vibrio cholerae		
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae, Vibrio.		
REFERENCE	1		
AUTHORS	Carlin, N. and Lebens, M.R.		
TITLE	Method of producing thy a<->strains of vibrio cholerae, such		
JOURNAL	strains and their use		
FEATURES	Patent: WO 961634-A 1 02-DEC-1999;		
source	CARLIN NILS (SB); SBL VACCIN AB. (SB); LEBENS MICHAEL R (SB)		
ORIGIN	location/Qualifiers		
	1..2909		
	/organism="Vibrio cholerae"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:566"		
Query Match	100.0%;	Score 2909;	DB 6; Length 2909;
Best Local Similarity	100.0%;	Pred. No. 0;	
Matches 2909;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;

QY 1 GAGAGTTTGTATGCTCAGGGTTATCTGACAGTTTCCCAATATTGACCCCGATTGTT 60  
 Db 1 GAGAAAGTTTGTATGCTCAGGGTTATCTGACAGTTTCCCAATATTGACCCCGATTGTT 60  
 QY 61 TTGCAATCGGCGCTTACGCGGTGCGGTGATAGCTTGAATATTGGTGGTTTCTTTT 120  
 Db 61 TTGCAATCGGCGCTTACGCGGTGCGGTGATAGCTTGAATATTGGTGGTTTCTTTT 120  
 QY 121 TGCATATGATGCGCAATCGCGAGCGGATCGCGGCGGAGTGGGTGGAGCGGTGAGA 180  
 Db 121 TGCATATGATGCGCAATCGCGAGCGGATCGCGGCGGAGTGGGTGGAGCGGTGAGA 180  
 QY 181 AGCTCTGACTTGTATTTCGCGGCTTTTAAAGTATGATCGGTGCGGAGTTGTTA 240  
 Db 181 AGCTCTGACTTGTATTTCGCGGCTTTTAAAGTATGATCGGTGCGGAGTTGTTA 240  
 QY 241 TGTGATCTTCTAATTTGATCTGTTCTTGTGTCGACCTCTTTATTTATCAAAAGTGTG 300  
 Db 241 TGTGATCTTCTAATTTGATCTGTTCTTGTGTCGACCTCTTTATTTATCAAAAGTGTG 300  
 QY 301 GACTGCGGCGATGCTCTTCCAGGCGGCTTATTGGGTGATCACCGGCAATGTTCTGTA 360  
 Db 301 GACTGCGGCGATGCTCTTCCAGGCGGCTTATTGGGTGATCACCGGCAATGTTCTGTA 360  
 QY 361 TGCAGGTAACCAACGACCTCTTGTGTGCGGATTTGTTGCCCTTTAGTGC 420  
 Db 361 TGCAGGTAACCAACGACCTCTTGTGTGCGGATTTGTTGCCCTTTAGTGC 420  
 QY 421 ATTGCGTTGGGATGGAGCTATCGGTAATGAAATGTAATGTAATGTAATGTAAT 480  
 Db 421 ATTGCGTTGGGATGGAGCTATCGGTAATGAAATGTAATGTAATGTAATGTAAT 480  
 QY 481 AAGCGATGCGCTTGGCTTTGTATTCCTATATGTTGGCCAGTCCGCGGCGATCTTC 540  
 Db 481 AAGCGATGCGCTTGGCTTTGTATTCCTATATGTTGGCCAGTCCGCGGCGATCTTC 540  
 QY 541 AAGCGATGCGCTTGGCTTTGTATTCCTATATGTTGGCCAGTCCGCGGCGATCTTC 540  
 Db 541 AAGCGATGCGCTTGGCTTTGTATTCCTATATGTTGGCCAGTCCGCGGCGATCTTC 540  
 QY 541 ACAGCTTTATGATTCGCTTGAAGGCGTGTCTGTTCTTATTTATTTATTTATTTAT 600  
 Db 541 ACAGCTTTATGATTCGCTTGAAGGCGTGTCTGTTCTTATTTATTTATTTATTTAT 600  
 QY 601 TGGTAAACCTCGTCCGCTAGGAGCGATTCGCGATCTGTTTAAAGCTGATACGTAAT 660  
 Db 601 TGGTAAACCTCGTCCGCTAGGAGCGATTCGCGATCTGTTTAAAGCTGATACGTAAT 660  
 QY 661 CCGCTTCCCTTGGATTAAGTCCGAGGAGCAATGTCAGTGGGTGGTGGTGGTGGT 720  
 Db 661 CCGCTTCCCTTGGATTAAGTCCGAGGAGCAATGTCAGTGGGTGGTGGTGGTGGT 720  
 QY 721 CATTTCAATGGGCAAACTCTCTCTTACCTATGATGATCATCGTATTTGATGATGAT 780  
 Db 721 CATTTCAATGGGCAAACTCTCTCTTACCTATGATGATCATCGTATTTGATGATGAT 780  
 QY 781 TTGGTCTTACAGCGGCTTGTATCAAGACCGGTGATAGCAAAATAGGATAGT 840  
 Db 781 TTGGTCTTACAGCGGCTTGTATCAAGACCGGTGATAGCAAAATAGGATAGT 840  
 QY 841 GAAACGATTTATGATCTTGTCAAGGAGTTCGATCAAGGTTGGTGGTGGTGGTGGT 900  
 Db 841 GAAACGATTTATGATCTTGTCAAGGAGTTCGATCAAGGTTGGTGGTGGTGGTGGT 900  
 QY 901 AGGAACGGGCAAGCGTGTGTTGACTGTGATTAATGCGATTTGACCTTACGATGTTG 960  
 Db 901 AGGAACGGGCAAGCGTGTGTTGACTGTGATTAATGCGATTTGACCTTACGATGTTG 960  
 QY 961 CAATCAGTTTCTTATGATGATCAACGCAAGAGTTTGGAAAGCTGCGGTAGCCGAT 1020  
 Db 961 CAATCAGTTTCTTATGATGATCAACGCAAGAGTTTGGAAAGCTGCGGTAGCCGAT 1020  
 QY 1021 GCTCGGCTATATTCGTTGATGATTAATGCGGAGATTTTGGCAATTTAGGTAACCAAA 1080  
 Db 1021 GCTCGGCTATATTCGTTGATGATTAATGCGGAGATTTTGGCAATTTAGGTAACCAAA 1080  
 QY 1081 CTGGATGCTAATGCAATTTAAACCAAGATGGCTCAACATCTTACGTAAGGTGA 1140

Db 1081 CTGGATGCTAATGCAATTTAAACCAAGATGGCTCAACATCTTACGTAAGGTGA 1140  
 QY 1141 GGAATGCAATGGAGCGCGTATGATGATCAGGGTATAGCTTGGGCTAAGCTGATGTTG 1200  
 Db 1141 GGAATGCAATGGAGCGCGTATGATGATCAGGGTATAGCTTGGGCTAAGCTGATGTTG 1200  
 QY 1201 TCATATTCAGCAATTTGAAAGATTTGATGATTTGAGCCGTGAGCTTGAATGACGAG 1260  
 Db 1201 TCATATTCAGCAATTTGAAAGATTTGATGATTTGAGCCGTGAGCTTGAATGACGAG 1260  
 QY 1261 TGAATTTCTTAACTTCAATTCGCGGTGAATTTCAATGCGGTGTTTGGCCCTTGAT 1320  
 Db 1261 TGAATTTCTTAACTTCAATTCGCGGTGAATTTCAATGCGGTGTTTGGCCCTTGAT 1320  
 QY 1321 GTACGCGCATATTTTCAATTTGCTGGGGGATACCTGTATCTCAACAGTACTCAGCTTC 1380  
 Db 1321 GTACGCGCATATTTTCAATTTGCTGGGGGATACCTGTATCTCAACAGTACTCAGCTTC 1380  
 QY 1381 ATGTATGCTGCTTGGGCTTGAATTTCAACATGATGAGGTTATGTTCTTGGCT 1440  
 Db 1381 ATGTATGCTGCTTGGGCTTGAATTTCAACATGATGAGGTTATGTTCTTGGCT 1440  
 QY 1441 GATGCGACATACAGGGAAGAACCGGCTTGGCGTATCAACAGATCTGATGCGCA 1500  
 Db 1441 GATGCGACATACAGGGAAGAACCGGCTTGGCGTATCAACAGATCTGATGCGCA 1500  
 QY 1501 CATTTACCAAGATCACTGGAATTTGATGCGGATGAGGCTTAAACGTTGAGCTTCC 1560  
 Db 1501 CATTTACCAAGATCACTGGAATTTGATGCGGATGAGGCTTAAACGTTGAGCTTCC 1560  
 QY 1561 AGCGCTCAGTTCCATCAATCCAAAGATTAAACATGCAAGATTGTAATGTAATGTAAT 1620  
 Db 1561 AGCGCTCAGTTCCATCAATCCAAAGATTAAACATGCAAGATTGTAATGTAATGTAAT 1620  
 QY 1621 CACTTGGATGATTTTGAAGTACCGGATACAGTACGATCTTATCAATCCCTT 1680  
 Db 1621 CACTTGGATGATTTTGAAGTACCGGATACAGTACGATCTTATCAATCCCTT 1680  
 QY 1681 TTCACTCAATCCGCTATTCAGGCGGTATGCTGATGGTGTATTAATTAATTAATTAAT 1740  
 Db 1681 TTCACTCAATCCGCTATTCAGGCGGTATGCTGATGGTGTATTAATTAATTAATTAAT 1740  
 QY 1741 CGAAGTGGGAGCTTTTATACAGATGATGCTTAAACCTTAAAGCGTTAGGGCAAG 1800  
 Db 1741 CGAAGTGGGAGCTTTTATACAGATGATGCTTAAACCTTAAAGCGTTAGGGCAAG 1800  
 QY 1801 AATGCTGCGGGATGAGCAAAACACCAATTAATGTAATCAACACCACTTTGCT 1860  
 Db 1801 AATGCTGCGGGATGAGCAAAACACCAATTAATGTAATCAACACCACTTTGCT 1860  
 QY 1861 CTTAACACCCCAAGTTGATGATGAGTCAACCTTATTAAGCAAGTGGCGTAAGAAAG 1920  
 Db 1861 CTTAACACCCCAAGTTGATGATGAGTCAACCTTATTAAGCAAGTGGCGTAAGAAAG 1920  
 QY 1921 AATACCGTAATCAAGACCGTAGCATCAAGTTAAAGCTTAAAGTCAACAGGCAATTTG 1980  
 Db 1921 AATACCGTAATCAAGACCGTAGCATCAAGTTAAAGCTTAAAGTCAACAGGCAATTTG 1980  
 QY 1981 CAGACCAACACCGCAAACTCAACAGACAGCGGTTGCGGAGCAAGCAGTAAATCA 2040  
 Db 1981 CAGACCAACACCGCAAACTCAACAGACAGCGGTTGCGGAGCAAGCAGTAAATCA 2040  
 QY 2041 AGTGCATATGTTGCACTTAAGTAAATGGGTGATTTCAACGCACTTTCAGACGCGCA 2100  
 Db 2041 AGTGCATATGTTGCACTTAAGTAAATGGGTGATTTCAACGCACTTTCAGACGCGCA 2100  
 QY 2101 GCCCAGAGAGGAACATTAAGCTGTTGTTGATGATGAAGTAAATTAATTAATTAAT 2160  
 Db 2101 GCCCAGAGAGGAACATTAAGCTGTTGTTGATGATGAAGTAAATTAATTAATTAAT 2160  
 QY 2161 CACTGTAACCTGAAGCAATACGTTAGTGGCTTGGCAATCGCATTTTGTAGAAATTC 2220

Db	2161	CACGTACCTGAAAGAAATACGCTGTAAGTGGGCGCTCGGCGCAATCGCATTTTGTAGAAATTTCC	2282
Qy	2221	ACGTGCGGGGCCCAACCATCTCTTTCATCAGTTTGCCCATCACCGTAATGGCGACGA	2286
Db	2221	ACGTGCGGGGCCCAACCATCTCTTTCATCAGTTTGCCCATCACCGTAATGGCGACGA	2286
Qy	2281	AATGGTCCGAATACCATAAGTAAAGTGGCGACACCGGAAGTATACCAATACCGGA	2340
Db	2281	AATGGTCCGAATACCATAAGTAAAGTGGCGACACCGGAAGTATACCAATACCGGA	2340
Qy	2341	AAGCTGGGTTTTCAGCCCTGTGATGACAGTTTGGTAAATCCGTTTGAATAAATCAAAC	2400
Db	2341	AAGCTGGGTTTTCAGCCCTGTGATGACAGTTTGGTAAATCCGTTTGAATAAATCAAAC	2400
Qy	2401	TTTTCATGCTCATATGCGCATGTCGACAGACGAGCGCAAAACGAGCCAGTGTGACCTTCTC	2460
Db	2401	TTTTCATGCTCATATGCGCATGTCGACAGACGAGCGCAAAACGAGCCAGTGTGACCTTCTC	2460
Qy	2461	TAAATATGCCAAATCATCTTTCAGAGGTAGGAAGATCGACACCGGGAAGAATTTGAAAAA	2520
Db	2461	TAAATATGCCAAATCATCTTTCAGAGGTAGGAAGATCGACACCGGGAAGAATTTGAAAAA	2520
Qy	2521	ATCGTGATGTTGTCGTCGTCGCAAGACGCGCAATCTTCTTTCACGCGCATATGGCC	2580
Db	2521	ATCGTGATGTTGTCGTCGTCGCAAGACGCGCAATCTTCTTTCACGCGCATATGGCC	2580
Qy	2581	AAGCTGACGAGATGTTGTCACAGTAAATCAATATTTGGCACCCATCACCATAGAAAT	2640
Db	2581	AAGCTGACGAGATGTTGTCACAGTAAATCAATATTTGGCACCCATCACCATAGAAAT	2640
Qy	2641	CGCGGTTTCACACCGGTAAACCAACGCGCAACGACCAATATTAAGATCACCGTCT	2700
Db	2641	CGCGGTTTCACACCGGTAAACCAACGCGCAACGACCAATATTAAGATCACCGTCT	2700
Qy	2701	TGAGGATTTGAATCAGTGGCGGTGTCACACTAAACCAATCATCTGCAATTTGGTGGGA	2760
Db	2701	TGAGGATTTGAATCAGTGGCGGTGTCACACTAAACCAATCATCTGCAATTTGGTGGGA	2760
Qy	2761	AGCAAAATTCAAATAGAACCTTTGGCTGATTCGCGCGGTCCCATTTTAAACCGCTGCCAG	2820
Db	2761	AGCAAAATTCAAATAGAACCTTTGGCTGATTCGCGCGGTCCCATTTTAAACCGCTGCCAG	2820
Qy	2821	CATCGCGACTGCAAGAAATGTAATATACGATGAAGCAAGTTTGGCCAACTGAGGCC	2880
Db	2821	CATCGCGACTGCAAGAAATGTAATATACGATGAAGCAAGTTTGGCCAACTGAGGCC	2880
Qy	2881	TTTGGTGGTCAAGCGAAATCGGCGCTGCAG	2909
Db	2881	TTTGGTGGTCAAGCGAAATCGGCGCTGCAG	2909
RESULT 3			
AE004153		9931 bp DNA linear BCT 10-OCT-2003	
LOCUS			
DEFINITION		Vibrio cholerae O1 biovar eltor str. N16961 chromosome I, section 61 of 251 of the complete chromosome.	
ACCESSION		AE004153 AE003852	
VERSION		AE004153.1 GI:9655103	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
Nature 406 (6795), 477-483 (2000)			

MEDLINE 20406833  
 PUBMED 10952301  
 REFERENCE 2 (bases 1 to 9931)  
 AUTHORS Heideberg,J.F., Eissen,J.A., Nelson,W.C., Clayton,R.A., Gwin,M.L., Dodson,R.J., Haft,D.H., Hickey,B.K., Peterson,J.D., Umayam,L.A., Gill,S.R., Nelson,K.E., Read,T.D., Tetteijn,H., Richardson,D., Ermolaeva,M.D., Vamathevan,J., Bae,S., Qin,H., Dracoli,I., Sellers,P., McDonald,L., Uterback,T., Fleishman,R.D., Nierman,W.C., White,O., Salzberg,S.L., Smith,H.O., Colwell,R.R., Mekalanos,J.J., Venter,J.C. and Fraser,C.M.  
 TITLE Submitted (14-JUN-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
 JOURNAL Direct Submission  
 FEATURES  
 SOURCE Location/Qualifiers  
 1. 9931  
 /organism="Vibrio cholerae O1 biovar eltor str. N16961"  
 /mol\_type="genomic DNA"  
 /strain="N16961"  
 /serotype="O1"  
 /db\_xref="taxon:243277"  
 /chromosome="1"  
 /note="biotype: El Tor"  
 288. .602  
 /gene="VC0666"  
 288. .602  
 /gene="VC0666"  
 /note="identified by Glimmer2; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAFP93831.1"  
 /db\_xref="GI:9655104"  
 /translation="MSSDPSLSIVIPPSKTTGENTROSIVQLAVLVRAELRREERVMRKRRSSSYHLPMMNPHLRDLGLETGRRPGLSLPDAVVAERVRRI RVLTAARI Pr"  
 complement (640. .1698)  
 /gene="VC0667"  
 complement (640. .1698)  
 /gene="VC0667"  
 /note="similar to PID:882727 GB:U00096 PID:1789199 PID:2342573; identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="oxidoreductase Tas, aldo/keto reductase family"  
 /protein\_id="AAFP93832.1"  
 /db\_xref="GI:9655105"  
 /translation="MALRGSAMQYTKLPHSSLSLSKICIGTTPRGSNOADAPOLDYALEGNVFTDAENTPPYPTTAQOTGKTEBPGNMLASGKREKVLATKVGAPRVPIYRDWMALDHRNIHQAVDSDLRLQTDYIDLYQLHWPQRQTFPQQLNVPYDKOEVSLEIETLEALNDLVRMGKRVYIGVSNSTPGWGSYVLRAEKELPRIVISIONPYNLNRSFEVLEAIIISHLEGVGLAYSPLAGLSGKYLNGARPAAGACTLHGFSSRYFTQGLIATEVYLAQOFGFDPAQOMALAFVNGRPVANSNIIGATVMEQLKSNLSDISNABLLQKIQIGITTYSNPCP"  
 complement (1891. .2556)  
 /gene="VC0668"  
 complement (1891. .2556)  
 /gene="VC0668"  
 /note="similar to GB:Y00113 SP:P06722 GB:U13631 PID:42065 PID:577425; identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="DNA mismatch repair protein Muth"  
 /protein\_id="AAFP93833.1"  
 /db\_xref="GI:9655106"  
 /translation="MKRAPTTOQELITRAOQIAGLSFAIADAGCTVPDDLRKDKGMVQGLVHMLGATAGSRPQDDFEHLGIELSIPISYTGKLETTFPCVADLTGVHGLTWBSHVAENKLSKVLMIIPVGRERIPLAERCVGSPPLMSPSPBEBAQIKAWEBLMEIVLGVAQIITAKHGEVILQIRKPAANGRALTEAYGANGRPIDALPRGFYLRPTQFAIILOR YXA"  
 complement (2943. .3035)  
 /gene="VC0669"  
 complement (2943. .3035)  
 /gene="VC0669"  
 /gene="VC0669"

```
/note="identified by Glimmer2; putative"
/codon_start=1
/translation="hypothetical protein"
/protein_id="AAP93834.1"
/db_xref="GI:9655107"
/translation="MRHTAHLPTFSVNNRVKKLGGCVKKGIYQQ"
3018..3164
/gene="VC0670"
3018..3164
/gene="VC0670"
/note="identified by Glimmer2; putative"
/codon_start=1
/translation="hypothetical protein"
/protein_id="AAP93835.1"
/db_xref="GI:9655108"
/translation="MRGMSHKIVNLELAQNLSFPTPTNKYRANHNSKFADLGMK
NQRN"
3124..3705
/gene="VC0671"
3124..3705
/gene="VC0671"
/note="similar to P1D:882723 GB:U00096 SP:Q46930
P1D:1789194; identified by sequence similarity; putative"
/codon_start=1
/translation="Murt/nudix family protein"
/protein_id="AAP93836.1"
/db_xref="GI:9655109"
/translation="MLRLGICGKINAIKEFTRLPYIDDGYRLNVGIVLCNNHGVF
WAKRYGSHWMPQPGGIDGESPEQAMFELYEYVGLTKGVKLVATSRHMLRYLPLK
RLVWRMSQVPCIGOKOKWFLRLCEDSKINMROSPEFGDMRWVSWYVPRQVVSF
KDYVRBAKERASIALMPKPKKVKGKENTHRG"
3708..5954
/gene="VC0672"
3708..5954
/gene="VC0672"
/note="similar to P1D:882722 GB:U00096 P1D:1255724
P1D:1789193; identified by sequence similarity; putative"
/codon_start=1
/translation="phosphoenolpyruvate-protein phosphotransferase"
/protein_id="AAP93837.1"
/db_xref="GI:9655110"
/translation="MLSQLRDIVEOVSKEDVYQALDIPVQTCRAMSECTCYLIAN
EEMHRLMATAQGLKFGDKHIDPDEGLVGVKRSABEINLAASKIPSKFYKQLG
EEVYSFLGTPIIYRQVGLVVOQKSLFSEMESEFLVTLAQLAVLMAQTQG
HRLSKQQAITGVAAISGVALGEFWMDDTQDLSEVLPASALDDPEGERLALAVEN
ALADPRRMKSLDGDINKRALIPDLFTLLNDPMLRKDLKQIQKGRADVALQVY
EVSNRPRMSDVYLRERKODIRBELGORLLFLLNTEINOPKIDPVLIVVELTASV
LAALPKDLIAVYSLEGAANSRAILSRLLGIPAYMGVSNIRDLNGKKGIYDGSKG
LFTSPSKTILNEIRALANERELSRVNBALREPACTLDGATIELNAGLSADTSLA
VNQGVGVGLYRTEISFLQHRFPSEBEQTQYRVHLNTPYRQRVMTLDDIGKPL
PLPILEBNDPFLGMRGIRFTLDHPDIFIIQLARMSAESGNLSILPMSIGAELO
DLKFTYQAYOEVSODPRVVMPOIGIMEVSMLYLPLADKIDFVSGENDTQY
LLAVRNARVADVYESRPAVVALKOIQOCATHOIPVCQGEIAGPICALILIG
LGYTLISMTNSVAKYKIVRISLAEQLQALFOALTQYSGEITNMALYIEKGFPA
GTVRAGKR"
5964..6758
/gene="VC0673"
5964..6758
/gene="VC0673"
/note="identified by Glimmer2; putative"
/codon_start=1
/translation="conserved hypothetical protein"
/protein_id="AAP93838.1"
/db_xref="GI:9655111"
/translation="MNIELALILLIGAVVGLAGLIGGGLVVPALLFLPYAGI
STELVQJALATSLATIIITSGSSALNHKLGNVDFVVKMLMPGVVGGFGSVVAB
WPSQYLPVFVGIYVCLISVQWYRSIRALQNHMPSAWTTTCGIGIGIATISLAIGG
GSLSYFLNRHGIEMKKAIGSSVCGFALAIAGMIGTILHGVENLPQYSIGYVLP
```

```
ALLAIAITSMUTRIGAKLATQMPFARIKRPFAIFLMCAVATMLPQ"
6787..7602
/gene="VC0674"
6787..7602
/gene="VC0674"
/note="similar to GP:3201564; identified by sequence
similarity; putative"
/codon_start=1
/translation="prolipo protein diacylglycerol transferase"
/protein_id="AAP93839.1"
/db_xref="GI:9655112"
/translation="MPQGYLQPNIDVLYSISGLAVRWGLMTLVGFLPAMILANR
ADRLAGVITAMREOVSDLLPAGFLGVVIGRVGIVLYFNEDLEADPLFLKWTGMSF
HGGLGIVITAMREOVSDLLPAGFLGVVIGRVGIVLYFNEDLEADPLFLKWTGMSF
WAPVFNCGPLPRHPSOLYFRLGIVLFLINMEIGKPRIGSVGLFLAGYGTFRF
LVYVREPDALGLRGGFISNGQILSLPMVITIGLIMVWSYKRLGYDRVAK"
7612..8463
/gene="VC0675"
7612..8463
/gene="VC0675"
/note="similar to GP:3114981; identified by sequence
similarity; putative"
/codon_start=1
/translation="chymotrypsin-like serine protease"
/protein_id="AAP93840.1"
/db_xref="GI:9655113"
/translation="MRQYLDLQORITVQGVVWENRGRCLTYINADLTVDVNNOF
PLVTRKSPKAAVABELGRTGYDNADPQOLGKTVDANANINQAMINPYRKGED
DMGRVYGVQAMAKPDGHIQDLKQIVDLSRGVDVGEILNTPDEEPMGLRIPC
MHSHFSLGTLGTLNYSORSDVPLGTFNFMVQVYFLAQLMAQITGKRPGLAAYKIV
NAHYODLEMRBQVLRKBPAPQFHNKIKITLQDLFTWVLDLFDVYQYFHD
IQYFSPV"
```

```
Query Match 98.7%; Score 2870.6; DB 1; Length 9931;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 2885; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 1 GAGAAAGTTGTTAAGCTCAGGGGTTATCTGAGTTCCGAATATGACCCCGATTGTT 60
Db 6774 GAGAAAGTTGTTAAGCTCAGGGGTTATCTGAGTTCCGAATATGACCCCGATTGTT 6833

QY 61 TTCGATCCGCCCTCAGCGGTTGCGCTGATAGCTTGAATGATTTGATGATGCTTTT 120
Db 6834 TTCGATCCGCCCTCAGCGGTTGCGCTGATAGCTTGAATGATTTGATGATGCTTTT 6893

QY 121 TGTGATGTTGTTGCGCAATCGCCGAGCGGATCGCGGCGAGTGTGACCGCTGAGCA 180
Db 6894 TGTGATGTTGTTGCGCAATCGCCGAGCGGATCGCGGCGAGTGTGACCGCTGAGCA 6953

QY 181 AGTCCTGACTGTTATTCGCGGCTTTTATGATGATGATGATGATGATGATGATGATGAT 240
Db 6954 AGTCCTGACTGTTATTCGCGGCTTTTATGATGATGATGATGATGATGATGATGATGAT 7013

QY 241 TGTGATGTTGTTGCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
Db 7014 TGTGATGTTGTTGCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 7073

QY 301 GACTGCGGCGATGTCCTTCACAGCGGCTTATGAGGTTGATCACCGCATGTTTGGTGA 360
Db 7074 GACTGCGGCGATGTCCTTCACAGCGGCTTATGAGGTTGATCACCGCATGTTTGGTGA 7133

QY 361 TGGCGGTAAAAACCAACGACCTCTTGTGATGATGATGATGATGATGATGATGATGATGAT 420
Db 7134 TGGCGGTAAAAACCAACGACCTCTTGTGATGATGATGATGATGATGATGATGATGATGAT 7193

QY 421 ATTGCGTTTGGGATGAGGACGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
Db 7194 ATTGCGTTTGGGATGAGGACGATGATGATGATGATGATGATGATGATGATGATGATGAT 7253

QY 481 AACGATGTCCTTGGGCTTTGATATTCCTAATGATGATGATGATGATGATGATGATGATGAT 540
```

Db 7254 AACGATGTCCTTGGGCTTTTGTATTCCTTAATGCTGCGCCACTGCGGCGCATCTTC 7313  
Qy 541 ACAGCTTATGAATTCGCTTAAAGAGCGTGTCTCTTATTTATTTATTTAT 600  
Db 7314 ACAGCTTATGAATTCGCTTAAAGAGCGTGTCTCTTATTTATTTATTTATTTAT 7373  
Qy 601 TGGTAAACCTGTCGCTAGGAGCGATATCCGATCTGTTTTAGCTGATACGATCATT 660  
Db 7374 TGGTAAACCTGTCGCTAGGAGCGATATCCGATCTGTTTTAGCTGATACGATCATT 7433  
Qy 661 CCGCTTCCTGTTGGAATACCTCCGTAGCCAGATACCTCAATGGGTCTGTTGGTCTT 720  
Db 7434 CCGCTTCCTGTTGGAATACCTCCGTAGCCAGATACCTCAATGGGTCTGTTGGTCTT 7493  
Qy 721 CATTTCAATGGGCAAAATCTCTCTTACCTATGATGATCATCGATTTTATGATGAT 780  
Db 7494 CATTTCAATGGGCAAAATCTCTCTTACCTATGATGATCATCGATTTTATGATGAT 7553  
Qy 781 TTGGTCTTACAGCGCGGTTGTATCAAGACCGTGTAGCAGCAAAATAGGGTATGAT 840  
Db 7554 TTGGTCTTACAGCGCGGTTGTATCAAGACCGTGTAGCAGCAAAATAGGGTATGAT 7613  
Qy 841 GAAACAGTATTTGATCTTTGTCAAGCGCATCTGATCAAGGTGTTGGTGAATA 900  
Db 7614 GAAACAGTATTTGATCTTTGTCAAGCGCATCTGATCAAGGTGTTGGTGAATA 7673  
Qy 901 AGCAAGCGGCAAGCGTGTGATCTGTATTAATGCGATTTAGCCTTACGATGAGGCA 960  
Db 7674 AGCAAGCGGCAAGCGTGTGATCTGTATTAATGCGATTTAGCCTTACGATGAGGCA 7733  
Qy 961 CAATCAGTTTCTCTAGTACTACAGCAGAGTTTTTGGAAAGCTGCTAGCCGAT 1020  
Db 7734 CAATCAGTTTCTCTAGTACTACAGCAGAGTTTTTGGAAAGCTGCTAGCCGAT 7793  
Qy 1021 GCTCGCTATATTCGTGTGATCAATATGCGCGGATTTTGGCAATTAGTACAAAC 1080  
Db 7794 GCTCGCTATATTCGTGTGATCAATATGCGCGGATTTTGGCAATTAGTACAAAC 7853  
Qy 1081 CTGGAGTGTATGCAATTTAAACCAAGATGAGCTCAACATCTTACCGTAAAGTGA 1140  
Db 7854 CTGGAGTGTATGCAATTTAAACCAAGATGAGCTCAACATCTTACCGTAAAGTGA 7913  
Qy 1141 GATGACATGAGGAGCGGTGTATGTGTTCAGGGTAGAGCTTGGGCTTAAAGCTGAT 1200  
Db 7914 GATGACATGAGGAGCGGTGTATGTGTTCAGGGTAGAGCTTGGGCTTAAAGCTGAT 7973  
Qy 1201 TCATATTTGACAGTGTGAAAAGATGTTGATGATTTGAGCGGTGATGACCGAG 1260  
Db 7974 TCATATTTGACAGTGTGAAAAGATGTTGATGATTTGAGCGGTGATGACCGAG 8033  
Qy 1261 TGAATTTCTTAACTTCTACAAATCCGGGTGAATTTCAATGAGGCTTGGCCCTTGAT 1320  
Db 8034 TGAATTTCTTAACTTCTACAAATCCGGGTGAATTTCAATGAGGCTTGGCCCTTGAT 8093  
Qy 1321 GPACAGCATATTTTCTATGCTGAGGAGTACCTTGTATCTCAACAGTACTAGGCTTC 1380  
Db 8094 GPACAGCATATTTTCTATGCTGAGGAGTACCTTGTATCTCAACAGTACTAGGCTTC 8153  
Qy 1381 ATGTGATGTGCTTGGGTTGAATTTCAATGATGAGCTTATGTTCTTGGCT 1440  
Db 8154 ATGTGATGTGCTTGGGTTGAATTTCAATGATGAGCTTATGTTCTTGGCT 8213  
Qy 1441 GATGACAGTATCAAGGGAAGAGCGGCTTGGGCTTACCAAGTCTCAATGCGCA 1500  
Db 8214 GATGACAGTATCAAGGGAAGAGCGGCTTGGGCTTACCAAGTCTCAATGCGCA 8273  
Qy 1501 CATTTACAGATCACTGAAATTTGAGCGGATGAGCTTAAACGTAAGCATTTCC 1560  
Db 8274 CATTTACAGATCACTGAAATTTGAGCGGATGAGCTTAAACGTAAGCATTTCC 8333  
Qy 1561 AGGCTCTAGTTTCAATATCAATCAAAAGATTTAAACATGACAGATTTGGAACTTGGGT 1620  
Db 8334 AGGCTCTAGTTTCAATATCAATCAAAAGATTTAAACATGACAGATTTGGAACTTGGGT 8393

Qy 1621 CACTTGTATGATTTTGAAGTACCGGATATCAGTTTCAAGATCTTATCAATACCGTT 1680  
Db 8394 CACTTGTATGATTTTGAAGTACCGGATATCAGTTTCAAGATCTTATCAATACCGTT 8453  
Qy 1681 TTCAGTCTAATCCGATATTCAGCGGATATGCTTGAATGGGTTTTATATTAAGAGCTCC 1740  
Db 8454 TTCAGTCTAATCCGATATTCAGCGGATATGCTTGAATGGGTTTTATATTAAGAGAGCTCC 8513  
Qy 1741 CGAAGTCCGAGCTTTTATTAAGATGATGATGATGATGATGATGATGATGATGATGAT 1800  
Db 8514 CGAAGTCCGAGCTTTTATTAAGATGATGATGATGATGATGATGATGATGATGATGAT 8573  
Qy 1801 AATGCTCGCGGATGACCAAAACACCAATATGATGATGATGATGATGATGATGATGAT 1860  
Db 8574 AATGCTCGCGGATGACCAAAACACCAATATGATGATGATGATGATGATGATGATGAT 8633  
Qy 1861 CTTACAGCCCAAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1920  
Db 8634 CTTACAGCCCAAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8693  
Qy 1921 AATACCGTAAATCAACCGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1980  
Db 8694 AATACCGTAAATCAACCGTATGATGATGATGATGATGATGATGATGATGATGATGAT 8753  
Qy 1981 CAGAGCAACACCGGCAACCTCACAGAGACGGGTTGGCGGAGAGAGAGATTAATAC 2040  
Db 8754 CAGAGCAACACCGGCAACCTCACAGAGACGGGTTGGCGGAGAGAGAGATTAATAC 8813  
Qy 2041 AGTCCCAATGTTGCACTTAAGGTAATGGTATGATGATGATGATGATGATGATGATGAT 2100  
Db 8814 AGTCCCAATGTTGCACTTAAGGTAATGGTATGATGATGATGATGATGATGATGATGAT 8873  
Qy 2101 GCGCAAGAGAGAACATTAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2160  
Db 8874 GCGCAAGAGAGAACATTAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 8933  
Qy 2161 CACTGATCTGAAGCAATACCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2220  
Db 8934 CACTGATCTGAAGCAATACCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 8993  
Qy 2221 AGGTGCGCGGCAACCATCAAACTTTCATCAAGTTTCCCATCAACCGTATGCGAGCA 2280  
Db 8994 AGGTGCGCGGCAACCATCAAACTTTCATCAAGTTTCCCATCAACCGTATGCGAGCA 9053  
Qy 2281 AATGTGCAATATCCCAATACGATGATGATGATGATGATGATGATGATGATGATGAT 2340  
Db 9054 AATGTGCGGATATCCCAATACGATGATGATGATGATGATGATGATGATGATGATGAT 9113  
Qy 2341 AAGCTGGGTTTCAAGCCCTGTATGATGATGATGATGATGATGATGATGATGATGAT 2400  
Db 9114 AAGCTGGGTTTCAAGCCCTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 9173  
Qy 2401 TTTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2460  
Db 9174 TTTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9233  
Qy 2461 TAAATGCAAAATCATATTTTGAAGTGAAGATCAGACCGCGAGAAATGTA 2520  
Db 9234 TAAATGCAAAATCATATTTTGAAGTGAAGATCAGACCGCGAGAAATGTA 9293  
Qy 2521 AATGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2580  
Db 9294 AATGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9353  
Qy 2581 AAGGCTGACGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2640  
Db 9354 AAGGCTGACGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9413  
Qy 2641 CCGGTTTCAACCGGTAACCAACCGGCAACGAGACCAATATATGAAAGTCAACGCTGT 2700  
Db 9414 CCGGTTTCAACCGGTAACCAACCGGCAACGAGACCAATATATGAAAGTCAACGCTGT 9473

Qy 2701 TGAGATTGAATCAGTCCGTTGGCCATTAACCAATCAATCCTGCAATTGGTGGA 2760  
Db 9474 TGAGATTGAATCAGTCCGTTGGCCATTAACCAATCAATCCTGCAATTGGTGGA 9533  
Qy 2761 AGCAATTCAATTAACCTTTGGCTTGATCGCGGTTGGCCATTTAAACCGCTGCCAC 2820  
Db 9534 AGCAATTCAATTAACCTTTGGCTTGATCGCGGTTGGCCATTTAAACCGCTGCCAC 9593  
Qy 2821 CATCGCATCTGCAGAAAGTAAATATACGATGAAGCAAGTTGGCCAACTGAGCC 2880  
Db 9594 CATCGCATCTGCAGAAAGTAAATATACGATGAAGCAAGTTGGCCAACTGAGCC 9653  
Qy 2881 TTTCGTGTCAGCGAATCGCGCTGCAG 2909  
Db 9654 TTTCGTGTCAGCGAATCGCGCTGCAG 9682

RESULT 4  
LOCUS VCTHYAG 1410 bp DNA linear BCT 01-NOV-2000  
DEFINITION Vibrio cholerae thya gene.  
ACCESSION Y17135.1 GI:3114980  
VERSION Y17135.1 GI:3114980  
KEYWORDS thya gene; thymidilate synthase.  
SOURCE Vibrio cholerae  
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae; Vibrio.  
REFERENCE 1 Valle, E., Ledon, T., Cedre, B., Campos, J., Valmaseda, T., Rodriguez, B., Garcia, L., Marreco, K., Benitez, J., Rodriguez, S. and Fando, R.  
TITLE Construction and characterization of a nonproliferative El Tor cholera vaccine candidate derived from strain 638  
INFECTION Immun. 68 (11), 6411-6418 (2000)  
JOURNAL MEDLINE 20490573  
PUBMED 11035753  
REFERENCES 2 (bases 1 to 1410)  
AUTHORS Benitez, J. A.  
TITLE Direct Submission  
JOURNAL Submitted (16-APR-1998) J. A. Benitez, Centro Nacional de Investigaciones, Cientificas, PO Box 6990, La Habana, CUBA  
FEATURES  
source location/Qualifiers  
1. 1410  
/organism="Vibrio cholerae"  
/mol\_type="genomic DNA"  
/strain="C7258"  
/db\_xref="taxon:666"  
289. 1140  
/gene="thya"  
289. 1140  
/gene="thya"  
/codon\_start=1  
/transl\_table=1  
/product="thymidilate synthase"  
/protein\_id="CAA76645.1"  
/db\_xref="gi:3114981"  
/db\_xref="GCA:066108"  
/db\_xref="SWISS-PROT:066108"  
/translation="MRQYLDLCQRIYDQGVWVENRTGRCLTVINADLTVDVGNNOF  
PLVTRKSFWMKAVALLEIGYIRGYNADFRQLGTWMDANLANLQAMLNPNYRGED  
DMGRVYGVQGRAMAKPDGGHIDOLKIVDDLSRGVDGEILNPNPGBPHMGCLRPG  
MYSHPFLIGDTLYLNTSDRSCDPLGINFMNVQYVPLAIAOITGKKPGIAVHKIV  
NAHYDQDELMKRDVQLKREPPAPQFHINPKIKTLQDLFTWTTLDDEVDVGYQYHPD  
IQYPSV"

ORIGIN  
Query Match 47.8%; Score 1390.8; DB 1; Length 1410;  
Best Local Similarity 99.1%; Pred. No. 0;  
Matches 1398; Conservative 0; Mismatches 12; Indels 0; Gaps 0;  
Qy 551 GAATTCGCTTGAAGAGCGCTGCTGTTCTTATTTCTTAATTTGTTATTTGTTAACT 610  
|||||

Db 1 GAATTCGCTTGAAGAGCGCTGCTTCTTATTTCTTAATTTGTTATTTGTTAACT 60  
Qy 611 CGTCGCTAGAGCAGCGATCCGACCTGTTTATAGCTGATACGGTACATTCGCTTCCT 670  
Db 61 CGTCGCTAGAGCAGCGATCCGACCTGTTTATAGCTGATACGGTACATTCGCTTCCT 120  
Qy 671 GTGAAATACGTCGCTGAGCCAGATGCTCAGTTGGGCTCTGTTTGGCTTCAATTCATG 730  
Db 121 GTGAAATACGTCGCTGAGCCAGATGCTCAGTTGGGCGCTGTTTGGCTTCAATTCATG 180  
Qy 731 GGGCAATCCCTCTCCTTACCTATGATGATCATCGGTAATTTGATGATGTTGGCTTAC 790  
Db 181 GGGCAATCCCTCTCCTTACCTATGATGATCATCGGTAATTTGATGATGTTGGCTTAC 240  
Qy 791 AAGCGCGTTTGTATCAAGACCGTGTAGCAGCAAAATAGGTAGTTAGTGAACAGTAT 850  
Db 241 AAGCGCGTTTGTATCAAGACCGTGTAGCAGCAAAATAGGTAGTTAGTGAACAGTAT 300  
Qy 851 TTAGATCTTTGTACAGCGCATGCTGATCAAGGTGTTGGTTGAAATGAACGAGGC 910  
Db 301 TTAGATCTTTGTACAGCGCATGCTGATCAAGGTGTTGGTTGAAATGAACGAGGC 360  
Qy 911 AAGCGTTGTTGACCTGTATTAATGCGATTTGACCTAAGATGTCGCAACATCAGTT 970  
Db 361 AAGCGTTGTTGACCTGTATTAATGCGATTTGACCTAAGATGTCGCAACATCAGTT 420  
Qy 971 CCTAGTGACTACACGCAAGTTTGGAAACCTGCGTAGCGAGTCTCGGCTAT 1030  
Db 421 CCTAGTGACTACACGCAAGTTTGGAAACCTGCGTAGCGAGTCTCGGCTAT 480  
Qy 1031 ATTGCTGTTACGATATATGCGCGGATTTTGGCAATTAGGTACCAAACTCGGATGCT 1090  
Db 481 ATTGCTGTTACGATATATGCGCGGATTTTGGCAATTAGGTACCAAACTCGGATGCT 540  
Qy 1091 AATGCAATTTAAACCAAGCATGCTGCAACAATCTTACCGTAAAGATGAGATGACATG 1150  
Db 541 AATGCAATTTAAACCAAGCATGCTGCAACAATCTTACCGTAAAGATGAGATGACATG 600  
Qy 1151 GAGCGCGTATAGTGTTCAAGGTTAGAGCTTGAGCTGATGATGATGATGATGATGATG 1210  
Db 601 GAGCGCGTATAGTGTTCAAGGTTAGAGCTTGAGCTGATGATGATGATGATGATGATG 660  
Qy 1211 CAGTTGAAAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1270  
Db 661 CAGTTGAAAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720  
Qy 1271 AACTTCTACATCCGGGTTGATTTCAATGAGGTTGTTGCCCTTGATGATGATGATGATG 1330  
Db 721 AACTTCTACATCCGGGTTGATTTCAATGAGGTTGTTGCCCTTGATGATGATGATGATG 780  
Qy 1331 CATTTTCAATGCTGGGGATATCCTTGTATCTCAACAGTACTAGCGTTCAATGATGATG 1390  
Db 781 CATTTTCAATGCTGGGGATATCCTTGTATCTCAACAGTACTAGCGTTCAATGATGATG 840  
Qy 1391 CCTTGGGGTTGAATTTCAACATGCTGAGGTTTATGTTCTTTGCGCTGATGAGCAG 1450  
Db 841 CCTTGGGGTTGAATTTCAACATGCTGAGGTTTATGTTCTTTGCGCTGATGAGCAG 900  
Qy 1451 ATCACAGGAAAGAGCCGGGCTTGCGCTGATCAAGATGCTGATGATGATGATGATGATG 1510  
Db 901 ATCACAGGAAAGAGCCGGGCTTGCGCTGATCAAGATGCTGATGATGATGATGATGATG 960  
Qy 1511 GATCAACTGAAATGATGCTGAGTGTGAGCTGATGAGCTGATGAGCTGATGAGCTGATG 1570  
Db 961 GATCAACTGAAATGATGCTGAGTGTGAGCTGATGAGCTGATGAGCTGATGAGCTGATG 1020  
Qy 1571 TTCCATATCAATCCAAAGATTTAAACACTGACGATTTGAAACTTTGGGTCACTTTGAT 1630  
Db 1021 TTCCATATCAATCCAAAGATTTAAACACTGACGATTTGAAACTTTGGGTCACTTTGAT 1080  
Qy 1631 GATTTTGAAGTCAAGGATATCAGTTTCAAGATCTTATTTCAATACCGTTTCACTTCA 1690  
Db 1081 GATTTTGAAGTCAAGGATATCAGTTTCAAGATCTTATTTCAATACCGTTTCACTTCA 1140

QY 1691 TCCCGTATTGAGCGGTATGCTTGTATGCTTTTATATATAAAGCTCCGGAAGTCCG 1750  
 Db 1141 TCCCGTATTGAGCGGTATGCTTGTATGCTTTTATATATAAAGCTCCGGAAGTCCG 1200  
 QY 1751 GAGCTTTTATATACAGATGATGCTTTAAAGCGTTAAGCGGTATGCGCAAGATGCTCCG 1810  
 Db 1201 GAGCTTTTATATACAGATGATGCTTTAAAGCGTTAAGCGGTATGCGCAAGATGCTCCG 1260  
 QY 1811 GGGATGACGACAAACACCCCAATAAGTAATCTACACCAACATTTTGTCTTACAAAGCC 1870  
 Db 1261 GGGATGACGACAAACACCCCAATAAGTAATCTACACCAACATTTTGTCTTACAAAGCC 1320  
 QY 1871 CAAGTTGAGATGAGCTCAGACACCTTTAATAGGAGCTTCCGTAAGAAAGAAATCCGTA 1930  
 Db 1321 CAAGTTGAGATGAGCTCAGACACCTTTAATAGGAGCTTCCGTAAGAAAGAAATCCGTA 1380  
 QY 1931 ATCAAGACCGTAGCCATCAAGTTAAAGCTT 1960  
 Db 1381 ATCAAGACCGTAGCCATCAAGTTAAAGCTT 1410

RESULT 5  
 AX009483 1222 bp DNA linear PAT 06-SEP-2000

LOCUS AX009483  
 DEFINITION Sequence 3 from Patent WO961634.  
 ACCESSION AX009483  
 VERSION AX009483.1 GI:9996768

KEYWORDS  
 SOURCE Vibrio cholerae  
 ORGANISM Vibrio cholerae  
 Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae; Vibrio.

REFERENCE 1  
 Carlin, N. and Lebens, M.R.  
 Method of producing chyl a<->strains of vibrio cholerae, such strains and their use  
 Patent: WO 961634-A 3 02-DEC-1999;  
 CARLIN NILS (SE); SBL VACCIN AB (SE); LEBENS MICHAEL R (SE)

JOURNAL  
 FEATURES  
 source 1..1222  
 /organism="Vibrio cholerae"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:666"

ORIGIN  
 Query Match 42.0%; Score 1222; DB 6; Length 1222;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1222; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1688 TAATCCCGTATCAGCGGTATGCTTGTATGCTTTTATATAAAGCTCCGGAAGT 1747  
 Db 1 TAATCCCGTATCAGCGGTATGCTTGTATGCTTTTATATAAAGCTCCGGAAGT 60

QY 1748 CGGAGCTTTTATACAGATGATGCTTTAAAGCGTTAAGCGGTATGCGCAAGATGCTG 1807  
 Db 61 CGGAGCTTTTATACAGATGATGCTTTAAAGCGTTAAGCGGTATGCGCAAGATGCTG 120

QY 1808 CCGGGAGTACGACAAACACCCCAATAAGTAATCTACACCAACATTTTGTCTTACAA 1867  
 Db 121 CCGGGAGTACGACAAACACCCCAATAAGTAATCTACACCAACATTTTGTCTTACAA 180

QY 1868 GCCCAAGTTGAGATGAGCTCAGACACCTTTAATAGGAGCTTCCGTAAGAAAGAAATCCG 1927  
 Db 181 GCCCAAGTTGAGATGAGCTCAGACACCTTTAATAGGAGCTTCCGTAAGAAAGAAATCCG 240

QY 1928 TAAATCAAGACCGTAGCCATCAAGTTAAAGCTTAAAGTGCACAGCGCAATTTGCAAGCA 1987  
 Db 241 TAAATCAAGACCGTAGCCATCAAGTTAAAGCTTAAAGTGCACAGCGCAATTTGCAAGCA 300

QY 1988 AACACGGCAAACTACACGAGACAGCGGTGCGCGAGACAGACAGTAATACAGTGCCA 2047  
 Db 301 AACACGGCAAACTACACGAGACAGCGGTGCGCGAGACAGACAGTAATACAGTGCCA 360

QY 2048 ATGTTGCACTTAAGTAAATGGTAGATTTTACGCACTTTTCAGCACGCGCAGAGCCACG 2107  
 Db 361 ATGTTGCACTTAAGTAAATGGTAGATTTTACGCACTTTTCAGCACGCGCAGAGCCACG 420

QY 2108 AAGAGAACCTTAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2167  
 Db 421 AAGAGAACCTTAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480

QY 2168 CCGTAAGCAATACCGGTATGAGGAGCTGCGCAATCGCAATTTTGTATGATTTTCAAGTGG 2227  
 Db 481 CCGTAAGCAATACCGGTATGAGGAGCTGCGCAATCGCAATTTTGTATGATTTTCAAGTGG 540

QY 2228 CCGCAACCATCAAACTCTTCATCACTTTGCCATCAACCGTAATGCGCAGAAATGCTC 2287  
 Db 541 CCGCAACCATCAAACTCTTCATCACTTTGCCATCAACCGTAATGCGCAGAAATGCTC 600

QY 2288 GCAATACCAATACGATTAAGTGCACACCAACGAAAGTATTAACCAATACCGAAAGCTG 2347  
 Db 601 GCAATACCAATACGATTAAGTGCACACCAACGAAAGTATTAACCAATACCGAAAGCTG 660

QY 2348 GTTTCAGGCTGTGATGACAGCTTGTATGCTTGTATGATTAATCAAAACCTTCATG 2407  
 Db 661 GTTTCAGGCTGTGATGACAGCTTGTATGCTTGTATGATTAATCAAAACCTTCATG 720

QY 2408 CTCATATCGCAGTGCAGACAGAGCGCAACGAGCAGTGTGAGACTTTCTTAAATG 2467  
 Db 721 CTCATATCGCAGTGCAGACAGAGCGCAACGAGCAGTGTGAGACTTTCTTAAATG 780

QY 2468 CCAAAATCATTTCTAGAGTAGAAGATCAGCACCGCGAAGATTTGAAAAATCGTGG 2527  
 Db 781 CCAAAATCATTTCTAGAGTAGAAGATCAGCACCGCGAAGATTTGAAAAATCGTGG 840

QY 2528 ATGTTGCACTGCGCAAGCAACGCGCAAACTTTCTTTACAGCGCATATGCGCAAGCTG 2587  
 Db 841 ATGTTGCACTGCGCAAGCAACGCGCAAACTTTCTTTACAGCGCATATGCGCAAGCTG 900

QY 2588 ACGAGATTTGCTGACAGTAGTACCAATATTTGGCAACCATCAACCATCAATGAGATTCGGGT 2647  
 Db 901 ACGAGATTTGCTGACAGTAGTACCAATATTTGGCAACCATCAACCATCAATGAGATTCGGGT 960

QY 2648 TCAACCGGTAAACCAACCGGCAACGAGCAACCAATATTAAGTACCGTCTTGAGAT 2707  
 Db 961 TCAACCGGTAAACCAACCGGCAACGAGCAACCAATATTAAGTACCGTCTTGAGAT 1020

QY 2708 TCAATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2767  
 Db 1081 TCAATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140

QY 2828 ACTGCAAGAGTAGTAATATACAGCATGAAGCAAGTTTGGCCCAAGTATGAGCTTTCGTG 2887  
 Db 1141 ACTGCAAGAGTAGTAATATACAGCATGAAGCAAGTTTGGCCCAAGTATGAGCTTTCGTG 1200

QY 2888 GTACGGAATGCGGCTGCGAG 2909  
 Db 1201 GTACGGAATGCGGCTGCGAG 1222

RESULT 6  
 AY143429 1236 bp DNA linear BCT 28-SEP-2002  
 LOCUS AY143429  
 DEFINITION Vibrio cholerae mutant chymidilase synthetase (chylA) gene, complete cds.  
 ACCESSION AY143429  
 VERSION AY143429.1 GI:23343945

KEYWORDS  
 ORGANISM Vibrio cholerae  
 Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;

REFERENCE 1 (bases 1 to 1236)  
Vibronaceae; Vibrio.  
Xia,X.  
AUTHORS  
TITLE Direct Submission  
JOURNAL Submitted (20-AUG-2002) Microbiology, Institute for Infectious  
Disease Control and Prevention, Chinese Center for Disease Control  
and Prevention, P.O. Box 5, Changping, Beijing 102206, P.R. China  
FEATURES  
source  
1. 1236  
/organism="Vibrio cholerae"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:666"  
248..1099  
/gene="chxA"  
/note="MchxA"  
248..1099  
/gene="chya"  
/note="does not catalyze the conversion of dUMP into dTMP"  
/codon\_start=1  
/transl\_table=1  
/product="mutant chymidilate synthetase"  
/protein\_id="AA16904.1"  
/db\_xref="GI:23343946"  
/translation="MRQYLDLCQRIVDGQWVENRGRKCTVTINADLTVDVGNQF  
PLVTRKSFWKAVALBELGLRGYNADFRQLGRTKDANANLQAMLNPRKGD  
DMGRVGVQGRAMAPDGGHIDOLKRIVDLSRGVDGEILNPNRPFHMGCLRPC  
MSHPSLLGDTLYNLSRSCDPLMLNFMWQYVFLAAMAQITGKKPLGAYKIV  
NAHIYODLELMRDVQKREPAPQPHNPKIKTLQDLFWTLDDVDVGYQPHDP  
IOYPSV"  
ORIGIN  
Query Match 41.8%; Score 1216.8; DB 1; Length 1236;  
Best Local Similarity 99.0%; Pred. No. 0;  
Matches 1224; Conservative 0; Mismatches 12; Indels 0; Gaps 0;  
QY 592 TTGGTTATTTGTTAAACCTGTCCTGAGGAGCGGATTCGGAGCTGTTTAACTGAGTA 651  
DB 1 TTGGTTATTTGTTAAACCTGTCCTGAGGAGCGGATTCGGAGCTGTTTAACTGAGTA 60  
QY 652 CGGTACATCCGCTCTCTGTTGGAATAGCTCCGTAGACAGATGCTCAGTGGGCTGTT 711  
DB 61 CGGTACATCCGCTCTCTGTTGGAATAGCTCCGTAGACAGATGCTCAGTGGGCTGTT 120  
QY 712 TGTGGCTTCATTTCAATGGGCAAAATCCTCTCCTTACCTATGATGATCATCGTATTTT 771  
DB 121 TGTGGCTTCATTTCAATGGGCAAAATCCTCTCCTTACCTATGATGATCATCGTATTTT 180  
QY 772 GATGATGTTTGTGCTTACAGCGCGGTTTGTATCAAGCCGTGTAGCAGCAAAATAGGG 831  
DB 181 GATGATGTTTGTGCTTACAGCGCGGTTTGTATCAAGCCGTGTAGCAGCAAAATAGGG 240  
QY 832 TAGTTAGGTGAAGACATATTAGATTTTGTCAAGCGCATGTCATGATGATGTTGGGT 891  
DB 241 TAGTTAGGTGAAGACATATTAGATTTTGTCAAGCGCATGTCATGATGATGTTGGGT 300  
QY 892 TGAATAATGAACGACGCGGAGCGTGTGTGACTGTGATTAATGCGATTTGACCTACGA 951  
DB 301 TGAATAATGAACGACGCGGAGCGTGTGTGACTGTGATTAATGCGATTTGACCTACGA 360  
QY 952 TGTGGGCAACATCAGTTTCTCTAGTGACTACACGCAAGATTTTGGAAAAGCTCCGT 1011  
DB 361 TGTGGGCAACATCAGTTTCTCTAGTGACTACACGCAAGATTTTGGAAAAGCTCCGT 420  
QY 1012 AGCCGAGTTGCTCGGCTATATTCGTGTGTGAAGATTAATGCGGATTTTGGCAATTAG 1071  
DB 421 GGCCTGAGTTGCTCGGCTATATTCGTGTGTGAAGATTAATGCGGATTTTGGCAATTAG 480  
QY 1072 TAGCCAAACCTGGGATGTAATGCCAATTAAACCAAGATGGCTCAACATCTTACCG 1131  
DB 481 TAGCCAAACCTGGGATGTAATGCCAATTAAACCAAGATGGCTCAACATCTTACCG 540  
QY 1132 TAAAGGTGAGATGATGAGGAGCGCGTATGTTGTGAGGGTAGAGCTTGGGCTTAAGCC 1191

DB 541 TAAAGGTGAGATGATGAGGAGCGCGTATGTTGTGAGGGTAGAGCTTGGGCTTAAGCC 600  
QY 1192 TGATGTGTGATATTTGACAGCTTGAAGAAAGATTTGATGATTTGAGCCGTGGCGTTGA 1251  
DB 601 TGATGTGTGATATTTGACAGCTTGAAGAAAGATTTGATGATTTGAGCCGTGGCGTTGA 660  
QY 1252 TGACCGAGGTGAATTTCTTAATCTTACATTCGGGTGAATTTTCAATGGGGTGTTCGG 1311  
DB 661 TGACCGAGGTGAATTTCTTAATCTTACATTCGGGTGAATTTTCAATGGGGTGTTCGG 720  
QY 1312 CCTTGCATGTACAGCCATCATTTTTCATGTGCTGGGGGATACCTTGTATCTCAACAGTAC 1371  
DB 721 CCTTGCATGTACAGCCATCATTTTTCATGTGCTGGGGATACCTTGTATCTCAACAGTAC 780  
QY 1372 TCAGCGTTCAATGTATGTGCTTGGGCTGAATTTTCAATGTGTGACAGTTTATGTGT 1431  
DB 781 TCAGCGTTCAATGTATGTGCTTGGGCTGAATTTTCAATGTGTGACAGTTTATGTGT 840  
QY 1432 CTTTGCCTGATGTGACAGATCAACAGGAAAAGCGGGCTTGGCGTATCAACAGATCGT 1491  
DB 841 CTTTGCCTGATGTGACAGATCAACAGGAAAAGCGGGCTTGGCGTATCAACAGATCGT 900  
QY 1492 CAATGGCGACATTTTACCAAGATCAACTGCAATGTGATGCGGATGTGCAAGCTTAAACGTA 1551  
DB 901 CAATGGCGACATTTTACCAAGATCAACTGCAATGTGATGCGGATGTGCAAGCTTAAACGTA 960  
QY 1552 GCCATTCCACGCGCTCAGTTCCATATCATATCAAAAGATTAAACACTGACAGATTGGA 1611  
DB 961 GCCATTCCACGCGCTCAGTTCCATATCATATCAAAAGATTAAACACTGACAGATTGGA 1020  
QY 1612 AACTTGGGTCACTTTGATGATATTTTACGTCAACCGGATATCAAGTTCCAGATCTTATCA 1671  
DB 1021 AACTTGGGTCACTTTGATGATATTTTACGTCAACCGGATATCAAGTTCCAGATCTTATCA 1080  
QY 1672 ATACCCGTTTCACTGATATCCCGATATCAGCGGATATGCTGATGCGGTTTATATAA 1731  
DB 1081 ATACCCGTTTCACTGATATCCCGATATCAGCGGATATGCTGATGCGGTTTATATAA 1140  
QY 1732 AAAAGCTCCGCAAGGTGCGGAGCTTTTATACAGATGATGCTTAAACGTTAAGCGGT 1791  
DB 1141 AAAAGCTCCGCAAGGTGCGGAGCTTTTATACAGATGATGCTTAAACGTTAAGCGGT 1200  
QY 1792 TAGGGCAAGATGCTGCGGCGGATGACGACAAACAC 1827  
DB 1201 TAGGGCAAGATGCTGCGGCGGATGACGACAAACAC 1236  
RESULT 7  
VCAJ10968/c  
LOCUS VCAJ10968 1591 bp DNA linear BCT 21-SEP-1998  
DEFINITION Vibrio cholerae nptA gene.  
ACCESSION AJ010968  
VERSION AJ010968.1 GI:3646476  
KEYWORDS nptA gene; phosphate pump.  
SOURCE  
ORGANISM  
Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
Vibronaceae; Vibrio.  
REFERENCE  
1  
AUTHORS Lebens,M., Soderlund,L.O., Lundquist,P. and Carlén,N.I.A.  
TITLE A putative sodium dependent phosphate pump in Vibrio cholerae  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1591)  
AUTHORS Carlén,N.I.A.  
TITLE Direct Submission  
JOURNAL Submitted (17-SEP-1998) Carlén N.I.A., SBL Vaccin AB, Department of  
Molecular biology, FMB, Jv 30, Stockholm, S-105 21, SWEDEN  
COMMENT  
nptA gene is located inbetween the nhar gene (AJ002395) and the  
chya gene (AJ006514) both running in opposite directions.  
FEATURES  
source  
1. 1591  
/organism="Vibrio cholerae"

```

/mol_type="genomic DNA"
/strain="J51569"
/db_xref="taxon:666"
245..1588
/gene="nptA"
-35_signal
245..250
/gene="nptA"
-10_signal
267..272
/gene="nptA"
373..376
/gene="nptA"
383..1532
RBS
/gene="nptA"
CDS
/gene="nptA"
/codon_start=1
/cranal_table=11
/product="sodium dependent phosphate pump"
/protein_id="CAA09443.1"
/db_xref="GI:3646477"
/db_xref="QGA:087918"
/db_xref="SPTREMBL:087918"
/translation="MINOATSPAPISLTTLGRMANLAFMLYLLLLAVAMVSGFRK
ATGDOAKVLEFASHPILAGIMIGLVATPILIOSSVTSIIIVGLVAGGLPETAIPVM
GANITGVTVTVLVLGHMRCKEFPFASATIHDPNLLAVLIPLEMRGILEKV
SHVLVPLATGDSMKGFDPFKPITKPVITGLTQLSVLNGFVALIVGILATP
VALTWGKLMKSLMGRAREILONAIGRPLHGIAGFAVTVLVQSSSTTSLMVLV
GSGVLKREIYPTLGANIGTCITALLAATAVSGEFAVTVLQIALVHLSPNMAATVLI
YGIPLRELPKAEILISTWACKSKWVVSyllGVFVVIPIGSLALTA"
1550..1588
stem_loop
/gene="nptA"
/note="shared with thyA gene running in opposite
direction"
/function="transcription termination"

```

## ORIGIN

```

Query Match      40.8%; Score 1187; DB 1; Length 1591;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1723 TTATATAAAAAAGCTCCGAGAGTGGGAGCTTTTATATACAGATGATGCTTTAAAGC 1782
DB 1591 TTATATAAAAAAGCTCCGAGAGTGGGAGCTTTTATATACAGATGATGCTTTAAAGC 1532
QY 1783 TTAAGCGTTAGGCAAGATGCTGCGGGATGACACAAACACCCATAAGTAAT 1842
DB 1531 TTAAGCGTTAGGCAAGATGCTGCGGGATGACACAAACACCCATAAGTAAT 1472
QY 1843 CACCAACCACTTTGCTCTTCAAGGCCCAAGTTGAGATGATGACACCTTTAATAG 1902
DB 1471 CACCAACCACTTTGCTCTTCAAGGCCCAAGTTGAGATGATGACACCTTTAATAG 1412
QY 1903 CAGTTGGCTAAGAAAGAAATACCGTAATCAGACCGTATGCAATCAAGTTAAAGCTTAA 1962
DB 1411 CAGTTGGCTAAGAAAGAAATACCGTAATCAGACCGTATGCAATCAAGTTAAAGCTTAA 1352
QY 1963 GTGCACCAAGCAATTTGACAGCAAAACGCGAAACTCACGAGACAGCGGTTGGCG 2022
DB 1351 GTGCACCAAGCAATTTGACAGCAAAACGCGAAACTCACGAGACAGCGGTTGGCG 1292
QY 2023 GAGCAGAGCAGTAATACAAGTGCATGTTCCGACCTAAGTAATGGGTAGATTTCACG 2082
DB 1291 GAGCAGAGCAGTAATACAAGTGCATGTTCCGACCTAAGTAATGGGTAGATTTCACG 1232
QY 2083 CACTTCACAGACCCGACGAGCCACGAGAGAAACATTAAGCTGTGTGTGTCATGAAGA 2142
DB 1231 CACTTCACAGACCCGACGAGCCACGAGAGAAACATTAAGCTGTGTGTGTCATGAAGA 1172
QY 2143 TTGAACATAATACCGTAACCACTGATCCTGAAGCAATACCGTAGTGGGCTCGGCAAT 2202
DB 1171 TTGAACATAATACCGTAACCACTGATCCTGAAGCAATACCGTAGTGGGCTCGGCAAT 1112
QY 2203 CGCATTTTGTAGATTTCAGTGGCGGCGCAACATCAATCTTTCATCAGTTTGGCCAT 2262
DB 1111 CGCATTTTGTAGATTTCAGTGGCGGCGCAACATCAATCTTTCATCAGTTTGGCCAT 1052

```

```

QY 2263 CACCGTAATGCGCAGCAAAAATGTCGCAATACCAATACGATATAGTGCACACCGAA 2322
DB 1051 CACCGTAATGCGCAGCAAAAATGTCGCAATACCAATACGATATAGTGCACACCGAA 992
QY 2323 AGTATTAACCAATACCGAAAGCTGGGTTTCAGACCCCTGTGATGACAGTTTGGTAATCG 2382
DB 991 AGTATTAACCAATACCGAAAGCTGGGTTTCAGACCCCTGTGATGACAGTTTGGTAATCG 932
QY 2383 TTTGATTAATAATCAAACTTTTCAATGCTCAATTCGCAAGTCGAGAGAGCGCAACGAG 2442
DB 931 TTTGATTAATAATCAAACTTTTCAATGCTCAATTCGCAAGTCGAGAGAGCGCAACGAG 872
QY 2443 CCAGTGTAGACTTTCTCTTAAATGCCAAACATCTTTCTAGAGTAGAAGATACGAC 2502
DB 871 CCAGTGTAGACTTTCTCTTAAATGCCAAACATCTTTCTAGAGTAGAAGATACGAC 812
QY 2503 CCGGAGAAATGAAAAAATCGTGATGTCGTCGAGACCTGCGAAGACCGGCAAACTCTTC 2562
DB 811 CCGGAGAAATGAAAAAATCGTGATGTCGTCGAGACCTGCGAAGACCGGCAAACTCTTC 752
QY 2563 TTTACAGCGCATATGCGCAAGCTGACGAGATTTGTCACAGTATGCAATATTTGC 2622
DB 751 TTTACAGCGCATATGCGCAAGCTGACGAGATTTGTCACAGTATGCAATATTTGC 692
QY 2623 ACCCATCAACATAGAAATGCGGTTTCAACCGGTAAACCAACGCGAAGACCAACAT 2682
DB 691 ACCCATCAACATAGAAATGCGGTTTCAACCGGTAAACCAACGCGAAGACCAACAT 632
QY 2683 AATGAAAGTCACCGTCTTGAAGATTTGAATCAAGTGCCTTGCATTAACATATCA 2742
DB 631 AATGAAAGTCACCGTCTTGAAGATTTGAATCAAGTGCCTTGCATTAACATATCA 572
QY 2743 TCCTGCAATTTGGTGGGAAGCAATTTCAATTAATGAACTTTGGCTTATGCCGTTGCCA 2802
DB 571 TCCTGCAATTTGGTGGGAAGCAATTTCAATTAATGAACTTTGGCTTATGCCGTTGCCA 512
QY 2803 TTTAAACCGCTGCGCAGCATGCGGCTGCAAGAAATGAGTAATACAGATGAAGACCA 2862
DB 511 TTTAAACCGCTGCGCAGCATGCGGCTGCAAGAAATGAGTAATACAGATGAAGACCA 452
QY 2863 GTTTCGCAACGTAAGGCTTTGTGTGTCAGCAAAATCGCGCTGCAG 2909
DB 451 GTTTCGCAACGTAAGGCTTTGTGTGTCAGCAAAATCGCGCTGCAG 405

RESULT 8
AP005074 298900 bp DNA linear BCT 05-MAR-2003
LOCUS Vibrio parahaemolyticus DNA, chromosome 1, complete sequence, 2/11.
DEFINITION AP005074 BR000031
ACCESSION AP005074.1 GI:28805287
VERSION
KEYWORDS
SOURCE
ORGANISM
Vibrio parahaemolyticus
Vibrio parahaemolyticus
Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
Vibrionaceae; Vibrio.
REFERENCE
1 Nasu, H., Iida, T., Sugahara, T., Yamachi, Y., Park, K. S., Yokoyama, K.,
Makino, K., Shinagawa, H. and Honda, T.
A filamentous phage associated with recent pandemic Vibrio
parahaemolyticus O3:k6 strains
J. Clin. Microbiol. 38 (6), 2156-2161 (2000)
MEDLINE 20295086
PUBMED 10834969
REFERENCE
2 Makino, K., Oshima, K., Kurokawa, K., Yokoyama, K., Uda, T.,
Tagomori, K., Iijima, Y., Najima, M., Nakano M., Yamashita, A.,
Kubota, Y., Kimura, S., Yasunaga, T., Honda, T., Shinagawa, H.,
Hattori, M. and Iida, T.
Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
distinct from that of V cholerae
Lancet 361 (9359), 743-749 (2003)
JOURNAL

```

MEDLINE 22508454  
PUBMED 12620739  
REFERENCE 3 (bases 1 to 298900)  
AUTHORS Oshima,K., Kurokawa,K., Makino,K., Yokoyama,K., Yasunaga,T.,  
Honda,T., Shingawa,H., Hattori,M. and Iida,T.  
TITLE Direct Submission  
JOURNAL Submitted (09-Apr-2002) Ken Kurokawa, Osaka University, Genome  
Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,  
Japan (E-mail:ken@gen-info.osaka-u.ac.jp,  
URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,  
Fax:81-6-6879-2047)  
COMMENT Genome project  
This clone was isolated from a patient presenting with acute  
gastroenteritis.  
FEATURES  
source location/Qualifiers  
1..298900  
/organism="Hibrio parahaemolyticus"  
/mol\_type="genomic DNA"  
/strain="O3:K6"  
/sub\_strain="RIMD 2210633"  
/db\_xref="taxon:670"  
/chromosome="1"  
115..315  
/gene="VP0305"  
115..315  
/gene="VP0305"  
/note="similar to PIR:S56442 percent identity 49 in 63 aa"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAC58568.1"  
/db\_xref="GI:28805288"  
/translation="MLRPAVYRPLQVAFVKTLLFKGQFTIVGVSGFCFQDNGKVLPLD  
VKDQRLAVFKEVNOAISLSA"  
complement(399..1037)  
/gene="VP0306"  
complement(399..1037)  
/gene="VP0306"  
/note="similar to GB:AAF95689.1 (AE004324) percent  
identity 74 in 208 aa"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="peptide methionine sulfoxide reductase"  
/protein\_id="BAC58569.1"  
/db\_xref="GI:28805289"  
/translation="MLNKQTLISIDALPGRQPMQIDCHFWNOSSLTAPLAHQOO  
ILGWCFCWGAERLFWOLDGVSTSVGAGFTPNPTEVCTGKGTGTEVVRVDE  
RVLISLAQLAVFWEKHDPQGRQGNLGTYSALITYSDQDEIADRSKIYCOAL  
BAELRSTITTEIVPAGPYFAETVHQYLAKNPDGVCIGTGVCFPPLQO"  
1205..2917  
/gene="VP0307"  
1205..2917  
/gene="VP0307"  
/note="similar to GB:AAF95689.1 (AE004323) percent  
identity 66 in 571 aa"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="BAC58570.1"  
/db\_xref="GI:28805290"  
/translation="MIKRALPIIVGLLVVSGSAVADVSLKLGINDGALNDNKVILSS  
IPEKYSTSLRQARLDOSTITBALNALGYTHAKTISISEGNDLIVIHKLPIYKIK  
VMDVVISGEAKDEEPTNLIAKSPKVRGILNAGYDLSKSGIRNALAQGFENDFK  
LNKLIVIPELNANVRLHYEDSGIRYHFGVEITGSOIWMENRVESMRPEIGPYLSD  
VGEYQNLSTNDMPSSVYEPDLSKLEDERELPIQSLAPAKNOIEGTIGSTDTGV  
RGTLLKKKPMWASRGHSPNTALSKPQGTITAGYKPILDVTLREYVLOLQGLKHLND  
RDTESLNSLAVERHMTLDGKWHKTVYRHLIENFSQGLQDDGVOPLRGATFESTRY  
RGSSPMWGDKOSTVETGDPALSLSTRVLRLLGSSWIRGIGENHRLFRLEGGANI  
TEEFKLSPSLRFFAGDNNIRGYEBSISPVDEGALTGAAYTLLSTLEYOYRYGN  
WMAATFYDIDGAFNDTPWKSGAGVIRWASPVGVSDFPAWGDLDEKNNERRIHFSL

gene GPEL"  
2914..6681  
/gene="VP0308"  
2914..6681  
/gene="VP0308"  
/note="similar to GB:AAF95688.1 (AE004323) percent  
identity 54 in 1258 aa"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="BAC58571.1"  
/db\_xref="GI:28805291"  
/translation="MIKRMKRYMISLSLIGLLITTFVATVLFTHPGKLAIWGA  
EKALPOLKVENIQSLEPFRPALQVSVDESLSHIDAVERTALAINRCFDPKVCVD  
ELALQGVFMPELPPASEEPEETPLRSISTVPLIFVNVSEFNDINLVQNDIW  
OTFSTALSMOGRDLVIAPTALKDINVLAPAPAEAPAEAEPEKSTDTPODILPEVW  
IPLVEARLDIHNPKLAGEPVPVYVNHGIVARAGRVYKTELMPEVGLSTQ  
VTLSDYPIKQADALYKQADAKQKSLSSGSGDLSNATTSLEYQATIKGDIQ  
LKTQLPPVSLKNVQAQWPLEGQSDYGVSVPSLAKQSLDGYEVALFETKSGKDIQV  
DVALNGKGTLEQIDLESIVVVTLGELSGKVMAMWAPINWQADLNKNIOPLQWKE  
AEGDISGSLSTSGQWQVSLPKLDIDGIRGYPINVEQOLEASDKKQGEDIDQ  
LTTQGLASHGNOLSAKGIKIDQIINDVENPDPAPKSVDDLAKQKQKALBSGSK  
EPDINDLALNQIMQOANVETTLTGGDVTPLPAPKXNLSTANNITYDIDKXSDA  
LEVSGDEKTLQDLDVSVSLTETBGTFKQKPEMTWDGALRRLTSSQGWMSIQ  
KSTAVKXNIDKQIANVOAHCWLQAKSSVCLTEDI SVKLTQAKLAINDFNDQKPL  
PPEKTLQSVVAAQFAKWAPEKKEPVTLSVMPRGVOEALQEPISKWMSFSPKAL  
AKDLKDEWLPDVNDGDLGKVSILNVSSEKPTIDKVALSTPLDLFLALGEYSI  
FKAMNTDLASGVDMHPKVGQPLIDOMKLGQVTPIDINSQGVINFGKHQADLDA  
GITTPDGKLEITGPADMRDLODMRTARVPKELKUMPMWVKI KVEPMDITDVTPL  
AKVGNLNLPMGRIVIELEPPSAVGSVDVITLNKDLPQVDEVAMFPNVETDINIK  
GDPPOLAFGLKGLKGLKSLANTQDKPEFIVGEVNIYDGSYRSFGQDILVEGKILMN  
GDADQPYVSIKAIINPNTODDVIAGVAVGEPADPSVLEIFSEPMQDANLSTLNG  
ODIDGESGNAMTTLLIGTSLAKSGVVGIEGAFSGVODLQDTPAGSGDSQVTSVSG  
ILPGLQYXGVGINSIGSEFTVRYRLMQDYLBAVSGVDSAVDLLYQEPD"  
6726..7073  
/gene="VP0309"  
6726..7073  
/gene="VP0309"  
/note="similar to GB:AAF95687.1 (AE004322) percent  
identity 64 in 115 aa"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="BAC58572.1"  
/db\_xref="GI:28805292"  
/translation="MQHLIFRYGTLRGEYVNHHTLSSNQPLGHETDAQSFYNVGSY  
PALSBGQSIQGEVYLIDDDTLVALDKLEDPVYRBSIATPFGQAMITYLYQTEOL  
TEETASGMCQRV"  
7170..7532  
/gene="VP0310"  
7170..7532  
/gene="VP0310"  
/note="unknown"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAC58573.1"  
/db\_xref="GI:28805293"  
/translation="MRVLIIGAMMMLISACAQTLTPASTVYTDMAQFGKQALDGLRE  
LSQRIAKLDGTNATPELIMAYONGVEQGEYCEGQSAVMLGVVAGPYPFGICDDVDP  
FPHQDYDGRSSAGAP"  
complement(7607..8137)  
/gene="VP0311"  
complement(7607..8137)  
/gene="VP0311"  
/note="similar to GB:AAF95686.1 (AE004322) percent  
identity 90 in 176 aa"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11







QY 1208 GACCAGTTGAAAAAGATTGTTGATGATTTGAGCCGCTGGCTTGATGACCGAGTGAAATT 1267  
 DB 214417 GACCAATTGCGTAATAATTGTCATGACCTGACGCGCGGAGTGAAGATCGCGGAGATTC 214358  
 QY 1268 CTTAACTTCTAACAATCCGGGTGAATTTCAATGGGGTGTTCGCCCTTGCATGTACAGC 1327  
 DB 214357 CTCAATTTCTCAACACCCAGGTAAATTTCAATAGGGTGTTCGCCCTTGTATGTACAGT 214298  
 QY 1328 CATCATTTTTCATGCTGGGGGATACCTGTATCTCAACAGTACTCAGCGTTCATGTAT 1387  
 DB 214297 CATCATCTCTCATTTGCTGGGTATACGTTATACCTCAATAGCACACAGCTTCGTTGAC 214238  
 QY 1388 GTGCCCTTGGGGTGAATTTCAACATGTGACAGTTTATGTTCCTTGGCGTGAAGCA 1447  
 DB 214237 GTGCCACGTGGGCTGAATTTCAACATGTGTCAGGTGATGTCTTGGCGCATGTGCA 214178  
 QY 1448 CAGATCAACAGGAAAAAGCCGGGCTTGGGTATCACAAGATGTCAATAGCCGACATTTAC 1507  
 DB 214177 CAGATTTACGGGCAAAAAACCGGGCCAGCATTTCCATAGATTTGTAATGCAATATCTAT 214118  
 QY 1508 CAAGATCACTGCAATTTGATGGCGATGTGACGCTAAACGAGCCATTCACGCGCT 1567  
 DB 214117 CAAGATCACTGCACTGATGCTGATGTGCACTCAACGCAACCGTTACAGCGCA 214058  
 QY 1568 CAGTTCATATCATTCCAAAAGATTAAACACTGACAGATTGGAACCTTGGGTCACTTG 1627  
 DB 214057 AGCTTCCACATCAACCTGAGATCAAAATCTTTGGAAGATCTGGAACCTTGGGTCACTTG 213998  
 QY 1628 GATGATTTTGAAGTCAACCGGATATCATGTTCCACGATCCTATTCATACCCGTTTCAAGTC 1687  
 DB 213997 GATGATTTCTGGGTGAAGGGTATGAACATCATGACCCCAATCCAGTACCATTCCTGGTA 213938  
 QY 1688 TAA 1690  
 DB 213937 TAA 213935

RESULT 10  
 LOCUS AP005332 249150 bp DNA linear BCT 04-DEC-2003  
 DEFINITION Vibrio vulnificus Y016 DNA, chromosome I, complete genome, section 3/14.  
 ACCESSION AP005332 BA000037  
 VERSION AP005332.1 GI:37197435  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 . Vibrio vulnificus Y016  
 Vibrio vulnificus Y016  
 Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 Vibrionaceae; Vibrio.  
 REFERENCE  
 AUTHORS  
 1 Chen, C.Y., Wu, K.M., Chang, Y.C., Chang, C.H., Tsai, H.C., Liao, T.L.,  
 Liu, Y.M., Chen, H.J., Shen, A.B., Li, J.C., Su, T.L., Shao, C.P.,  
 Lee, C.T., Hor, L.I. and Tsai, S.F.  
 TITLE  
 JOURNAL Comparative Genome Analysis of *Vibrio vulnificus*, a Marine Pathogen  
 REFERENCE Genome Res. 13, 2577-2587 (2003)  
 AUTHORS 2 (bases 1 to 249150)  
 JOURNAL  
 TITLE  
 DIRECT SUBMISSION  
 JOURNAL Submitted (29-May-2002) Shih Feng Tsai, National Health Research  
 Institutes, Division of Molecular and Genomic Medicine; 128,  
 Yen-Chu-Yuan Road, Sec 2, Taipei, Taiwan 115, Republic of China  
 (E-mail: pepsai@nhri.org.tw, Tel: 886-2-8146-1041,  
 Fax: 886-2-2789-0484)  
 COMMENT  
 This sequence was determined by the Sequencing Core of the National  
 Yang-Ming University Genome Research Center (YMGRC;  
 http://genome.ym.edu.tw).  
 FEATURES  
 source  
 1. 249150  
 Location/Qualifiers  
 /organism="Vibrio vulnificus Y016"  
 /mol\_type="genomic DNA"  
 /strain="Y016"  
 /db\_xref="taxon:196600"  
 /chromosome="I"

gene 299..979  
 /gene="VV0511"  
 CDS 299..979  
 /gene="VV0511"  
 /note="identified by Glimmer2"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="BAC93275.1"  
 /db\_xref="GI:37197436"  
 /translation="MNVRSRKNNRLITTKTFLGRPIQTEHGPLVDYLEKMENTID  
 IALDEPYRLAIPVDLAPRKLKRNENSGNWTDLRLSQIDHSGKRVESGVRAH  
 PCKVRIAWREKSSINDHHLVLFNKDRPNMIGKTQIQENIGSPIVEKAVAGV  
 DYEEANGIVHPSKKNDSVIVIRLNCNSAFDADFEPKHISYLAKENKHFSGSR  
 RNFSKSHK"  
 865..888  
 /gene="rpt112"  
 865..888  
 /gene="rpt112"  
 /note="2 units of 12 bp in-tandem repeat"  
 /rpt\_type=tandem  
 /rpt\_unit="TGATGGTTGA"  
 1040..1708  
 /gene="VV0512"  
 1040..1708  
 /gene="VV0512"  
 /note="identified by GeneMark and Glimmer2"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="BAC93276.1"  
 /db\_xref="GI:37197437"  
 /translation="MTQLVNVAHQRAHTTDRTPENGKPLYPFDGLVLEYLEIDRV  
 LTELANKPRIVAAHINLPLSDGQDYLVEFAHPRLIESEINELCKDYDTRYR  
 OQVIVRYMTKASRSTSDHHLILFPDRIQNGTREGMRPLYYKVRKAMSVSE  
 AYRQPCSGLGISTNBSFELSDAEVFPLOANFPRIISRLAPITLNDHRKSG  
 CSHD"  
 complement(1286..1441)  
 /gene="VV0513"  
 complement(1286..1441)  
 /gene="VV0513"  
 /note="identified by Glimmer2"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAC93277.1"  
 /db\_xref="GI:37197438"  
 /translation="MKELPIKQDKMVIAGGLTRLRPLDIANNRLVISISGVVLT  
 ELISLRP"  
 1733..2368  
 /gene="VV0514"  
 1733..2368  
 /gene="VV0514"  
 /note="identified by GeneMark and Glimmer2"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="BAC93278.1"  
 /db\_xref="GI:37197439"  
 /translation="WPSFTHKKALLMANNRTYLPSTILDRTFEDYPLYYSGKYESA  
 LTVADVBEQALSQFDIALARLDLPEDHQDHLSTLNDYFNLKELKLDLSLFY  
 WRKEDKVSNSHNVMLFSDYOHGPAICMDKXNELVGHKTMQEAIEKRYGSGEAR  
 SLVFLNDRGNVGVTRCSKQVNIKNVTFHMSLSLAEMEKRYCFGSLS"  
 2436..2642  
 /gene="VV0515"  
 2436..2642  
 /gene="VV0515"  
 /note="identified by GeneMark and Glimmer2"  
 /codon\_start=1  
 /transl\_table=11  
 /product="predicted transcriptional regulator"  
 /protein\_id="BAC93279.1"  
 /db\_xref="GI:37197440"

```
gene /translacion="MMKRLRIKDVMSLTGLGRSTIYYFMADPETHDFPSVVLGRRAAVA  
WBSREEMMEORLALRDNPFQ"  
CDS /gene="VV0516"  
       .3162  
       2725..3162  
     /gene="VV0516"  
   /codon_start=1  
 /transl_table=11  
 /product="ribonuclease HI"  
 /protein_id="BAC93280.1"  
 /db_xref="GI:37197441"  
 /translation="MSYSIIYVDGAAPNNHGGCTRGIGIVMDEEDNEILVHEESFTINR  
KIDNALBELIALIEALEFYBEDDVIYSSSDICYKFNFNIIMDDMKDRCKRADKKFVNKA  
RHLMQQVDLSRKRYVEAKFAHSGBEGNERADLLAVEAALA"  
       .4423  
       3182..4423  
     /gene="VV0517"  
       .4423  
       3182..4423  
   /note="identified by GeneMark and Glimmer2"  
 /codon_start=1  
 /transl_table=11  
 /product="hypothetical protein"  
 /protein_id="BAC93281.1"  
 /db_xref="GI:37197442"  
 /translation="MGCTPGQAAILIHYHCWLREFELTFQCPRNELSLMLNKEFRMP  
VTIKGNPLSPDPVSASFGLPVSPPREKIKNLNASVNVSKLPTRTLDTNSKEBIIRPK  
SIRGFOMFLPPDEVSAFLPDGVSHKSIDLSPFWSSHKLKGVOQLSGTSSELNM  
GLLVSYIDRCVCAYOROLAFIODADINNENHELISGYVLVPDLFTLSEBALVKAR  
NMUKSESLEASTEQAMLCPCYDWFLSMVNYDIPEMDDKHAETBPVNAOXGQ  
LGQVNRERPTFGKFSGLIKEQTNOTTAOLSEYVLPUSETERNELLKRDOCAAF  
KPMRSKSIIPSVSEALSTEFPANNDAGROGADDLIYLFMRALDKVGSRLLPIYTKVGS  
RSLPDIYTISRVRFOHYAP"  
         4310..  
         4381  
       /gene="rptIT13"  
         4310..  
         4381  
     /gene="rptIT13"  
   /note="2 units of 36 bp in-tandem repeat"  
 /rpt_type=tandem  
 /rpt_unit="AAGCGGTAGTCGCCTCTACTCATATCTACAGC"  
 complement(4874..5101)  
 /gene="VV0518"  
 complement(4874..5101)  
 /note="identified by GeneMark and Glimmer2"  
 /gene="VV0518"  
   /codon_start=1  
 /transl_table=11  
 /product="conserved hypothetical protein"  
 /protein_id="BAC93282.1"  
 /db_xref="GI:37197443"  
 /translation="MNKSTLLIATALVFSTSSAFAGTTLAGNPADSQA VVVSVGKCT  
OKYMADKAODAGKESRTTVPGCEALKEK"  
 complement(5149..5508)  
 /gene="GV0519"  
 complement(5149..5508)  
 /note="identified by GeneMark and Glimmer2"  
 /gene="VV0519"  
   /codon_start=1  
 /transl_table=11  
 /product="conserved hypothetical protein"  
 /protein_id="BAC93283.1"  
 /db_xref="GI:37197444"  
 /translation="NNTLRIRTMALLLPSITACSBODAOTVADVESISITPHHGNG  
TIYPDIOLAITSHTTGAVIRINGCNRLRAGTTSLSRVOLKPBEITYGISNNCTUWS  
VKEIPTSSVGVFIYTYF"  
 complement(5532..5861)  
 /gene="VV0520"  
 complement(5532..5861)  
 /note="identified by Glimmer2"  
 /codon_start=1  
 /transl_table=11  
 /product="conserved hypothetical protein"  
 /protein_id="BAC93284.1"
```

gene	/db_xref="GI:37197445" /translation="WALPTIREMSDEGROVLNAVHTLAIHNRMLKTSQVLOGEKALI EPLGSVCCDMLLFLRSQGITRIPLALIDEASVSTESFTYPTDFNERPREHA EIRASQI" complement (5965 . 6273) /gene="VV0521" complement (5965 . 6273) /gene="VV0521" /note="identified by Genemark and Glimmer2" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="BAC93285.1" /db_xref="GI:37197446" /translation="MSNKTISIEGLSALHITMLIPORHWITVRELOOQALDIHRTT RSIKRYLDIIVDVNECDMSMPHYRYKTSQQLKLNQEMALLATDQGVLTASS "
gene	complement (6412 . 6588) /gene="VV0522" complement (6412 . 6588)
CDS	

Query Match	35.4%	Score 1028.6;	DB 1;	Length 249150;
Best Local Similarity	75.7%;	Pred. No. 2.8e-307;		
Matches 1274;	Conservative 0;	Mismatches 409;	Indels 0;	Gaps 0;

QY	8	TTTGTATAGCCTCAGGGTTATCTGCAATTTCCCAATATTGACCCCGATTTGTTTGATC	67
Db	175147	TTCAATTAATGTCTCAGGGTTACTGTGCTTTTCTTAATATGACCCCGCTTTTCTCGAATA	175206
QY	68	GGCCCTCAGACGGTCCGCTGGTATGAGCTTATGTTATGTTGAGGTTTCTTTTGTCTATG	127
Db	175207	GGACCAATCTCAATTCGTTGGTACGGGTGATGATCTAATTTGGTTTTTGTGGGATG	175266
QY	128	TGTTTGCCCAATCGCCGAGCCGATCGCGGGCAGTGTGGAACGCGTAGCAAGTCTCT	187
Db	175267	TGGCTCGCAATCGCCGAGCAGACAACACGGGCAAGTGGCTGACGCGAGACGAAGTTTGG	175326
QY	188	GACTTGTATTTCGCCGCTTTTATGTTATGATTCGGTGGCCGAGTTGTTATGATC	247
Db	175327	GATTTGCTGTTGCTCGAATCTCTAGGCGTTGTGCTGGGGTGTCGCAATTTGGGATCGTCTG	175386
QY	248	TTCTACAAATTTGATCTGTCTCTGCTGACCTCTTATTTAATTCAAAGTGGACATGGC	307
Db	175387	TTCTAATACCTTCGATCTCTCTCTCGCCGATTCGAAATTTACTCTTTTAAAGCTGGACAGT	175446
QY	308	GGCATATCTCTTCACAGCGCGCTTATTTGAGTGTATCAACCGCATGTTCTGTATGCGCGT	367
Db	175447	GGCATATCTTTTCACAGCGCGCTGTGCTGGGTGTCAATTAAGCATATGCTCTGTATCGCGAAG	175506
QY	368	AAAAACAACGCACTTCTTGTGTGTGCGCATTTTGTGTCCCTTTAGTGCATTCGCT	427
Db	175507	AAAAATGCGCAACTTTTGTGGCGTGGCGGACCTTCGTGTCCGTTAGTGCATTTGGT	175566
QY	428	TTGGGGATGGGACGATACGTATCTTTATGATATAGGAACCTTGGGAGACGATACGAT	487
Db	175567	TTAGGGTAGGGCGCTTAGGCACTTATAGAACGCGGAACCTTGGGCAAGAGTACCGAT	175626
QY	488	GTGCTTGGGCTTTTGTATCTCCCTATGTGTGGCCACATCGCGGCATCTTCACAGCTT	547
Db	175627	GTGCTTGGGCGATGTGTGTCCCAACCGCGCGGACCACTACACAGCACCTTCGCAATTG	175686
QY	548	TATGATTCGCTTAGAAGACGTGTCTCTTCTTATTCCTAATTTGTTTATTTGTTAA	607
Db	175687	TATGAGATGGCAATTAGAAGGGGTGTGTCTCTTATCTCAACGTGTTTATTCGCAAA	175746
QY	608	CCTGTCGGCTAGGACGAGTATCCGGAATGTTTATAGCTGATACGGTATCATTCGCGCTTC	667
Db	175747	CCAAACCTTTAGATCCGTAATCTGTATTATCTTGTCTGGTTATAGTACATTCGCTTTC	175806
QY	668	CTTGGGAATACGTCCGTGAGCCGAATCTCAGTTGGGCTCTGTTGGTGGCTTCAATTTCA	727
Db	175807	CTCGTTGAATACGTCCGTGAGCCGAATCCGAATCCGAATCGGATCTGTGTTGGTTCATTTCA	175866

```

QY 728 ATGGGCAATCTCTCTTACCTATGCTGATCATCGGATTTTGTATGATGTTGGTCT 787
   |||
Db 175867 ATGGGACAGATCTTGCTGCACCGATGATTTTGGTGGTGGCCCTTGTGGTGGCT 175926
   |||
QY 788 TACAGCGCGTTTGATCAAGCGGTGAGCAGAAAATAGGTTAGTTAGTGAACAG 847
   |||
Db 175927 TATTAACGTGGCCACTATCAAGCAAGTAACTGTAAGTTAGATGAGTGAACAA 175986
   |||
QY 848 TATTTAGATCTTTGTCAGCGCATCGTCATCAAGGTGTTGGGTGAAATAGCAAG 907
   |||
Db 175987 TATTTAGATGTTGCGCGTCATCGTATGTAAGGGCACTGGGTTGAAATAGACTACC 176046
   |||
QY 908 GCGAAGCGTTGTTGATCTGTGATTAATGCCGATTTGACTTACGATGTGGGCAACATCAG 967
   |||
Db 176047 GGTAAAGCGTGCCTTACTGTGTATCAACGCGGATTTAACTCGATGATGTCCTCAACATCAG 176106
   |||
QY 968 TTTCCTAGTACTACACGCAAGATTTTGGAAAGCTGGCGGTGCGAGTGGCTGGC 1027
   |||
Db 176107 TTTCCTTTGGTTTACACGCGTAAAGTTTCTGAAAGCGCGCGGTTGCAAGTTGCTTGGC 176166
   |||
QY 1028 TATATTCGTGTTACGATATGCGCGGATTTTTCGCAATTTAGTACCAAACTCGGAT 1087
   |||
Db 176167 TATATTCGAGGCTATGACATGCGGAAGATTTTTCGAAAGTTAGTACGAAACCTGGAT 176226
   |||
QY 1088 GCTAATGCCAATTTAAACCAAGCATGCTCAACAATCTTACCGTTAAGGTGAGATGAC 1147
   |||
Db 176227 GCGAAGCGCAATCTCAATGATGCTTGTGTAATTAACCCATATTCGAAAGCGCATGATGAC 176286
   |||
QY 1148 ATGGGACGGTGTATGAGGTGATGAGGTGATGAGGTGATGAGGTGATGAGGTGAT 1207
   |||
Db 176287 ATGGGCGGTGTCTACGCGCTACAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAT 176346
   |||
QY 1208 GACGATGAAAGATTTGATGATGATTTGAGCCGCGGTGATGATGATGATGATGAT 1267
   |||
Db 176347 GACCAATGCGTAATATGTCATGATGATGATGATGATGATGATGATGATGATGAT 176406
   |||
QY 1268 CTTAATCTTCAATCCGGGTAAATTTACATGAGGTGTTTCCGCTTGATGATGATGAT 1327
   |||
Db 176407 CTTAATCTTCAATCCGGGTAAATTTATGAGGTGTTTCCGCTTGATGATGATGAT 176466
   |||
QY 1328 CATCATTTTTCATGCTGGGGATGATGATGATGATGATGATGATGATGATGATGAT 1387
   |||
Db 176467 CATCATTTTTCATGCTGGGGATGATGATGATGATGATGATGATGATGATGATGAT 176526
   |||
QY 1388 GTGCCCTGGGGTGAATTTCAACATGAGGTGATGATGATGATGATGATGATGATGAT 1447
   |||
Db 176527 GTACCGCTGGGGTGAATTTCAACATGAGGTGATGATGATGATGATGATGATGATGAT 176586
   |||
QY 1448 CAGATCAAGGAAAAAGCGGGCTTGGCGTATCAACAGATGCTCATGCGCATTTAC 1507
   |||
Db 176587 CAGATCAAGGAAAAAGCGGGCTTGGCGTATCAACAGATGCTCATGCGCATTTAC 176646
   |||
QY 1508 CAAATCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1567
   |||
Db 176647 GAAAGTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 176706
   |||
QY 1568 CAGTTCATATCAATCCAAAGATTTAAACATGAGGTGATGATGATGATGATGATGAT 1627
   |||
Db 176707 AGCTTCATATCAATCCCTGATGATGATGATGATGATGATGATGATGATGATGAT 176766
   |||
QY 1628 GATGATTTGACGTCAAGGATATCATGATGATGATGATGATGATGATGATGATGAT 1687
   |||
Db 176767 GATGATTTCTGGGTGAAAGGATATGAAATCATGATGATGATGATGATGATGAT 176826
   |||
QY 1688 TAA 1690
   |||
Db 176827 TAA 176829

```

```

ACCESSION AX020429
VERSION AX020429.1 GI:10044145
KEYWORDS
SOURCE
ORGANISM
REFERENCE
  1 Campos,G.J., Ledon,P.T., Rodriguez,G.B., Benitez,R.J., Pando,C.R.,
    Silva,C.A. and Valle,D.E.
  TITLE Vibrio cholerae vaccine candidates and method of their constructing
  JOURNAL Patent: WO 9935271-A 1 15-Jul-1999;
  CAMPOS GOMEZ JAVIER (CU); LEDON PEREZ TRILENA YAMILLE (CU); RODRIGUEZ
  GONZALEZ BORIS LOTIS (CU); BENITEZ ROBLES JORGE ANTONIO (CU); CENT
  NAC INVESTIG SCIENT (CU); PANDO CALADA RAFAEL ALFREDO (CU); SILVA
  CARRERA ANISIA JUANA (CU); VALLE DIAZ EDGAR (CU)
  location/Qualifiers
  1..852
  /organism="Vibrio cholerae"
  /mol_type="unassigned DNA"
  /db_xref="taxon:666"
  1..852
  /note="unnamed protein product; thya nucleotide sequence"
  /codon_start=1
  /transl_table=1
  /protein_id="CAC07770.1"
  /db_xref="GI:10044146"
  /db_xref="EMBL:CA07770"
  /translation="MROYLDCRILVQGVNENRTKRCCLIVINADLTVDGNNOF
  PLATRKSPKRAVAALVGLGYRGYDANAPROGCTGTGTAANANANLNNPYRKED
  DMGRVYVQGNAMAKPBGHIDQJIKYVDLSRVDRLGELINYNNGEPRMGLRRC
  MYSHFSLIDTLINSTORSQDVPLGNTFMVVYPLAMAGITGKRPGLAYHKIV
  NAHYVDOLLEMRDVOLKRPBPAPQPHINKIKITLDLETWTLDFDVTGYQFHPD
  IQYFYSV"
  misc_difference 1..3
  /note="Xaa= Methionine"
  ORIGIN
  Query Match 28.9%; Score 841.4; DB 6; Length 852;
  Best Local Similarity 99.3%; Pred. No. 1.7e-249;
  Matches 845; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
  QY 840 TGAACAGTATTTAGATTTTGTGACGCCATCGTCATCAAGGTGTTGGTGAATAAG 899
  Db 2 TGAGCAGTATTTAGATTTTGTGACGCCATCGTCATCAAGGTGTTGGTGAATAAG 61
  QY 900 AACGAAGGGGAAGGCTTGTGACGTGATTANAGCCGATTGACATGAGGCA 959
  Db 62 AACGAAGGGGAAGGCTTGTGACGTGATTANAGCCGATTGACATGAGGCA 121
  QY 960 ACAATCAGTTTCTTACTGACTACACGCAAGATTTTGGAAAGCTGCGTAGCCAGT 1019
  Db 122 ACAATCAGTTTCTTACTGACTACACGCAAGATTTTGGAAAGCTGCGTAGCCAGT 181
  QY 1020 TGCTCGCTATATTTGTGTTACGATTAATGCGGCGGATTTTCCGCAATTAGATCAAAA 1079
  Db 182 TGCTCGCTATATTTGTGTTACGATTAATGCGGCGGATTTTCCGCAATTAGATCAAAA 241
  QY 1080 CCGTGGAGCTTAAGGCCAATTTAAACGAAGATGAGCTCAAAATCTTAACGGTAAAGGTG 1139
  Db 242 CCGTGGAGCTTAAGGCCAATTTAAACGAAGATGAGCTCAAAATCTTAACGGTAAAGGTG 301
  QY 1140 AGATGATAGGGAAGCGCTGATGAGTTCAGGGTAGAGTGGGCTAAGCTGATGGTG 1199
  Db 302 AGATGATAGGGAAGCGCTGATGAGTTCAGGGTAGAGTGGGCTAAGCTGATGGTG 361
  QY 1200 GTCATATTTGACCAAGTTGAAAAAGATTTGTGATGATTTGAGCCGTGATGACCGAG 1259
  Db 362 GTCATATTTGACCAAGTTGAAAAAGATTTGTGATGATTTGAGCCGTGATGACCGAG 421
  QY 1260 GTGAATTTCTTAATCTTCACAATCCGGGTGAATTTCAACAGGGGTGTTGGCCCTTGCA 1319
  Db 422 GTGAATTTCTTAATCTTCACAATCCGGGTGAATTTCAACAGGGGTGTTGGCCCTTGCA 481

```

RESULT 11  
 LOCUS AX020429 852 bp DNA linear PAT 07-SEP-2000  
 DEFINITION Sequence 1 from Patent WO9935271.

QY 1320 TGTACGACCATCTTTTTCATGCTGGGGGATACCTTGATCTCAACAGACTCAGGTT 1379  
 DB 482 TGTACGACCATCTTTTTCATGCTGGGGGATACCTTGATCTCAACAGACTCAGGTT 541  
 QY 1380 CATGTATGTGCTTGGGGTGAATTTCAACATGCTGAGTTTATGTGCTTGGC 1439  
 DB 542 CATGTATGTGCTTGGGGTGAATTTCAACATGCTGAGTTTATGTGCTTGGC 601  
 QY 1440 TGTATGCAACATCAACAGGAAAAAGCCGGGCTTGGCGTATCAACATGCTCAATGCGC 1499  
 DB 602 TGTATGCAACATCAACAGGAAAAAGCCGGGCTTGGCGTATCAACATGCTCAATGCGC 661  
 QY 1500 ACATTTACCAAGATCAACTGCAATGATGCGGATGCGAGTGAATAACGAGCCATTC 1559  
 DB 662 ACATTTACCAAGATCAACTGCAATGATGCGGATGCGAGTGAATAACGAGCCATTC 721  
 QY 1560 CAGCGCTCAGTTCATATCAATCAAGATTAACACATGCAAGATTTGAAACTTGGG 1619  
 DB 722 CAGCGCTCAGTTCATATCAATCAAGATTAACACATGCAAGATTTGAAACTTGGG 781  
 QY 1620 TCACCTTGGATGATTTTGAAGTCAACCGGATATGATTCACGATCCTATTCATCCGT 1679  
 DB 782 TCACCTTGGATGATTTTGAAGTCAACCGGATATGATTCACGATCCTATTCATCCGT 841  
 QY 1680 TTTCACTTAA 1690  
 DB 842 TTTCACTTAA 852  
 RESULT 12  
 AX009482 838 bp DNA linear PAT 06-SEP-2000  
 LOCUS AX009482  
 DEFINITION Sequence 2 from Patent WO961634.  
 ACCESSION AX009482  
 VERSION AX009482.1 GI:9996767  
 KEYWORDS  
 SOURCE Vibrio cholerae  
 ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae; Vibrio.  
 REFERENCE 1  
 AUTHORS Carlin, N. and Lebens, M.R.  
 TITLE Method of producing thy a<->strains of vibrio cholerae, such strains and their use  
 JOURNAL Patent: WO 961634-A 2 02-DEC-1999;  
 CARLIN NILS (SE); SBL VACCIN AB (SE); LEBENS MICHAEL R (SE)  
 FEATURES  
 source 1. 838  
 /organism="Vibrio cholerae"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:666"  
 ORIGIN  
 Query Match 28.8%; Score 838; DB 6; Length 838;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-248; Indels 0; Gaps 0;  
 Matches 838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GAGAAAGTTTGTATGCTCAGGAGTATGTCAGTTTCCCAATATGACCCCGTATGTT 60  
 DB 1 GAGAAAGTTTGTATGCTCAGGAGTATGTCAGTTTCCCAATATGACCCCGTATGTT 60  
 QY 61 TTGCATGCGGCTTCTAGCGGCTGCTGCTGATAGCTTGAATTTGGTGGCTTCTTT 120  
 DB 61 TTGCATGCGGCTTCTAGCGGCTGCTGCTGATAGCTTGAATTTGGTGGCTTCTTT 120  
 QY 61 TTGCATGCGGCTTCTAGCGGCTGCTGCTGATAGCTTGAATTTGGTGGCTTCTTT 120  
 DB 61 TTGCATGCGGCTTCTAGCGGCTGCTGCTGATAGCTTGAATTTGGTGGCTTCTTT 120  
 QY 121 TGTATGATGTTGGCAATCGCGAGCGATCGCGGCGAGTGTGGAACGCGTGAACA 180  
 DB 121 TGTATGATGTTGGCAATCGCGAGCGATCGCGGCGAGTGTGGAACGCGTGAACA 180  
 QY 181 AGTCTCTGACTTGTATTCGCGGCTTTTGAAGTGTAGTATCGATCGGCGAGTGTGTTA 240  
 DB 181 AGTCTCTGACTTGTATTCGCGGCTTTTGAAGTGTAGTATCGATCGGCGAGTGTGTTA 240

QY 241 TGTGATCTTCAATTTTATGATCTGTTCTTGTGACCCCTTTATTTATCAAGTGTG 300  
 DB 241 TGTGATCTTCAATTTTATGATCTGTTCTTGTGACCCCTTTATTTATCAAGTGTG 300  
 QY 301 GACTGCGGATGTCCTTCCACGCGGCTTATTTGGTGTATCACCGCATGTTCTGTA 360  
 DB 301 GACTGCGGATGTCCTTCCACGCGGCTTATTTGGTGTATCACCGCATGTTCTGTA 360  
 QY 361 TGTGCGTAAAAACCAACGACCTTCTTGTGTGCGCAATTTTGTGCCCCCTTATGTC 420  
 DB 361 TGTGCGTAAAAACCAACGACCTTCTTGTGTGCGCAATTTTGTGCCCCCTTATGTC 420  
 QY 421 ATTGCGTTTGGGATGGAAGGATGATGCTTATGATAGTGAATTTGGGAGAGT 480  
 DB 421 ATTGCGTTTGGGATGGAAGGATGATGCTTATGATAGTGAATTTGGGAGAGT 480  
 QY 481 AACGATGCTGCTTGGGCTTTTGTATTCCTATATGATGAGCCGATCCGCGCATCTTC 540  
 DB 481 AACGATGCTGCTTGGGCTTTTGTATTCCTATATGATGAGCCGATCCGCGCATCTTC 540  
 QY 541 AAGCTTATGATATGCGCTTGAAGAGCGTGTCTGTTCTTATTTCTTAATGTTTAT 600  
 DB 541 AAGCTTATGATATGCGCTTGAAGAGCGTGTCTGTTCTTATTTCTTAATGTTTAT 600  
 QY 601 TGTAAACCTGCTGCGCTAGAGGAGGTATCCGAGCTTTTATAGCTGATAGGTAATT 660  
 DB 601 TGTAAACCTGCTGCGCTAGAGGAGGTATCCGAGCTTTTATAGCTGATAGGTAATT 660  
 QY 661 CCGCTTCTTGTGATAAGTCCGTAGAGCAAGTCTCAGTTGGTCTGTTGGTGGCTT 720  
 DB 661 CCGCTTCTTGTGATAAGTCCGTAGAGCAAGTCTCAGTTGGTCTGTTGGTGGCTT 720  
 QY 721 CATTTCAATGAGGCAATCTTCTTCTTACTTATGATGATCATCGTATTTGATGATGT 780  
 DB 721 CATTTCAATGAGGCAATCTTCTTCTTACTTATGATGATCATCGTATTTGATGATGT 780  
 QY 781 TTTGCTTCAACAGCGCGGTTTGTATCAAGACCGGTAGAGCAAAATGGGTAGTTAG 838  
 DB 781 TTTGCTTCAACAGCGCGGTTTGTATCAAGACCGGTAGAGCAAAATGGGTAGTTAG 838  
 RESULT 13  
 AE006043/c 10719 bp DNA linear BCT 08-MAR-2001  
 LOCUS AE006043  
 DEFINITION Pasteurella multocida PM70 section 10 of 204 of the complete genome.  
 ACCESSION AE006043 AE004439  
 VERSION AE006043.1 GI:12720282  
 KEYWORDS  
 SOURCE Pasteurella multocida  
 ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales; Pasteurellaceae; Pasteurella.  
 REFERENCE 1  
 AUTHORS May, B.J., Zhang, Q., Li, L.L., Paustian, M.L., Whitlam, T.S. and Kapur, V.  
 TITLE Complete genomic sequence of Pasteurella multocida, PM70  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (6), 3460-3465 (2001)  
 MEDLINE 21145866  
 PUBMED 11248100  
 REFERENCE 2 (bases 1 to 10719)  
 AUTHORS Zhang, Q. and Kapur, V.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-OCT-2000) Department of Veterinary Pathobiology, University of Minnesota, 1971 Commonwealth Ave., St. Paul, MN 55108, USA  
 FEATURES  
 source 1. 10719  
 /organism="Pasteurella multocida"  
 /mol\_type="genomic DNA"  
 /strain="PM70"  
 /db\_xref="taxon:747"

gene 225. .1076  
/gene="PM0074"  
CDS 225. .1076  
/gene="PM0074"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAK02158.1"  
/db\_xref="GI:12720283"  
/translation="MKSSESTSCWSPTRKAIQIGEDVYVYKATKQIYSPFSAIPAG  
APIATAPVYATQTQNGNDAPKGLTKYGVIFSGVIMVVCSELFTSSSTMATVH  
VSGRISTQMLKNMLVIGGNVGLFTVLLTFAGQIMANGQMLTILTKAQKH  
HTWTEAFALGIFCNIMVCIAMWTTAGTKLLDKAVIMLPIISFVASGFESHVAMEM  
IPMGMMIAHFAISPEFWQAVSLNPAQADLDWTFVVKMLIPVLLGNIVGALCVALLQ  
WYINRPH"  
gene 1156. .3486  
/gene="pf1b"  
/note="synonym: PM0075"  
CDS 1156. .3486  
/gene="pf1b"  
/codon\_start=1  
/transl\_table=11  
/product="pf1b"  
/protein\_id="AAK02159.1"  
/db\_xref="GI:12720284"  
/translation="MMSQMLNETQOKAWEGFTGDMQTEVNVVRDPIQKNYTPYEGBES  
FLADYEAATKLMVDYMERIKYENTTHEYDIDCDTPSTITSHAGYIDKLSLEKIVL  
OTDAPLEKAIMPRGGIMNWKSCXYRARELKEVEGQIFTEYKTHNGVDPVYPTDIL  
RRKSGVITGLPDAYGRRITIDYRYRMALYGDPIMKDKFQFTSLQKLRGEDIQA  
TIQREIEAQORALQKMKEMASTGYDISEPATVAHEVQNTYPAITLAAYSQNGAA  
MSFGVSTFLDIYERDLKAKRITQEOAQLIDHLVMKLRMVFRLRTEYDQLEFSGD  
MWATETLAGMGIDGRITLVTKNSFRILHTLYTWGPSPPEPMLITLMSEKLPBEGFRAAK  
VSIDTSVQYENDDLMRPFDONDYALACVSPMIVGKMQPFGARALIAKTLILYAIN  
GGVDKSGDGVGPKTDPIITSEYLDYDWTTRLDSPMDMLAKOYVTLINIIHMHKVA  
YEAALMALHDRVPRFTMACGIAGLSVAADSLSAITYAKYKPRGDIETKXKAGVYGI  
AADVAIDETISEYFQPGKNDNRVDEIACDLVERWKCIQKLGITRNATPTOSVLTIT  
SNVYVGKTKGTGTPDRSGAPPGPAPNPMHGDQKAVASLTSVAKLPFAAKDQISY  
TFEIVNALSGDYEAQKNLGLAGMDGYFHEATITIEGQHILNVNVMNRBMLDAMENPE  
KYPOLITRVSQYAVRNSLTKEQQDDVITRFTQAM"  
gene 3659. .5698  
/gene="est"  
/note="synonym: PM0076"  
CDS 3659. .5698  
/gene="est"  
/codon\_start=1  
/transl\_table=11  
/product="est"  
/protein\_id="AAK02160.1"  
/db\_xref="GI:12720285"  
/translation="MKINTLSFALPCSLSSAALAKDVVFGDSLDMGQTGMKKASY  
LKADGSYHLVDEYLAQAGKLLIPSTOGILNAVSGGVIYGAHNTRTAEOPHLLEK  
QINELYIAHPKKEALHILMAGNDLATVATATVKTTPBEKQAVYLASINTMAQTMQ  
QNGALQAGVNOITIAPTIPNVYTPPEFPDKLGBAAGIQAKASYGLIKOSDPVANFKA  
AAEQULVQPTKLEERERKRVQTLERAAEDFYNSRWFTQILSSAGINAKIATLTI  
KEYKTIVBQAQATAFLNASITSALNOVGNIIVRIDTGLADMWTRPAEYXLTITTT  
VCKKSTADAPKACKBEDQATASORLFEFSDPHAPKAMSDITLWLTQPKMGI  
LTOIVOOQTEALDPIRTESNSRRLQOOSAGOSQVILAAOKOGGSHLVGAQVFN  
QMOALIVASOKODYQOGLININTONRUTSTLRYDADMMWLSGLVONNTLLTDRV  
AITGSAHQQSARETAESEFSKGLPAGYEKFKALGIALIADLNKTKTVAGGEBORNG  
ITQWQPAARTKESVRSVGFDLRYQASWQPVYTRVWKENNDIPTIOAGINSSFM  
TVLPQSDONQWNIYVAGLHPKPISSGFHANLAFRRDLGRKNALSNITLLQAGIGFTP"  
gene 5862. .6602  
/gene="act"  
/note="synonym: PM0077"  
CDS 5862. .6602  
/gene="act"  
/codon\_start=1  
/transl\_table=11  
/product="act"  
/protein\_id="AAK02161.1"  
/db\_xref="GI:12720286"  
/translation="MAVVGRIHSYSBSGCTVDPGIRFILFMQGLMRQCYCHNRDITWD  
LHAGKEITVEELMKEVVYTRHFMNASGGGCVTASGSAVLQAEFVDMWFSACKKEGHT

gene CUDTNGFVRNVYDVIDELLDVTDLVLIDLKQJLNDKTHQNLICVFNKRTLEPQYIAYR  
NQPVWIRYVWVPGYTADDEDIHLIGHITIKMKNIKELVLLPYRHLGAHKWEAMEKYE  
LEEVNPTKESLEHIIKSLSEYGHIIKY"  
gene complement (6819. .7319)  
/gene="PM0078"  
CDS complement (6819. .7319)  
/gene="PM0078"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAK02162.1"  
/db\_xref="GI:12720287"  
/translation="MQSDMDKMKMRHRLADKABQJGEIPVGAVIDEGONLLGEGH  
NLISSESDPTMAHAIVALRQGGRLQYRLLNTTLVYLEPTCAGALHSRIKRLV  
FGASDYKTGAWSRFHFFDDYKONHALEITAGVLOEBCSOKISAFQKRREQKIAQO  
KSHLET"  
gene complement (7334. .8185)  
/gene="thya"  
/note="synonym: PM0079"  
CDS complement (7334. .8185)  
/gene="thya"  
/codon\_start=1  
/transl\_table=11  
/product="thya"  
/protein\_id="AAK02163.1"  
/db\_xref="GI:12720288"  
/translation="MKHYLELCQRIIVDEGVWENARTGKCLTVIDADLTYYDVAANO  
FPLITTRSYWAAIAEFLGYIRGYDNADEFKLGTKTDANAANSAMLNPHKRGTD  
DMGRVYGVGQAMRKPNGETVDOLRKIVNLRNGIDRGEIMTFEFGSEFLGCLRCP  
MHTYTPSLSGTLLVTSYQBSCDPILGNFQIOLVPRILAMQDITCKKGAKAHKII  
NAHYEDDLMKVQLRDRDFPLPOLHINPEIKTLEDLETWTTMDPQKTYGYOCHDA  
IKYPSV"  
gene complement (8182. .8994)  
/gene="tgt"  
/note="synonym: PM0080"  
CDS complement (8182. .8994)  
/gene="tgt"  
/codon\_start=1  
/transl\_table=11  
/product="tgt"  
/protein\_id="AAK02164.1"  
/db\_xref="GI:12720289"  
/translation="MQSYRFSFPQPDVYIFETIGPGLRWYGMILGLFPRMILAYR  
ANQYSGWTTDQVDSLLENFGWGFVLGRIQYVLYQDFLDPAFLFRWBGMSF  
HGGILGVITAMLITAKIQKRGFMQTADEFVALIFPGDMGRIGNF INDELMGRVTDVP  
WAVLEPSSGCVLPRBSQLYEAVLEIGVLFILNMYIKKPRIGATAGLFLIGYGIFFR  
IVERFRBDALQGLYFGQHISMGOILSTPMLIGAVIMLVAGSAGKKREIL"  
gene complement (9004. .9798)  
/gene="PM0081"  
CDS complement (9004. .9798)  
/gene="PM0081"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAK02165.1"  
/db\_xref="GI:12720290"  
/translation="MFTFELICLVAGAFAGLIGFGIGGLVTPVLVYLMPKXVGV  
PEAWMACALGTSFATVITGFSASORHKGANVMDVKYLAIVIMTTFISGFPIS  
GLPIDISARLPACTIVYAIKMPISVXKPYTKLITKASTIAGILGMASSAGIIG  
GGLVPIPLANSKGTDMKKAIGSSAVCGSLIAAGNLSPNIAQWKYBGLPEYMSGYVLP  
ALVGITLTSITSIKIGATVISILFVPIUKAFALMLILAINMLIK"  
gene complement (9801. .10397)  
/gene="PM0082"  
CDS complement (9801. .10397)  
/gene="PM0082"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAK02166.1"  
/db\_xref="GI:12720291"  
/translation="MIDPDGYPNVGIVICNSKQOVLAKRYGQNSMQPQGGIINDNE  
SABQAMRLELFEYGLSPKDVKILYISGMRLYKLPKRLIARYDSKSPVICIGKQKRFLL  
QVDEKXINNQSSKSPFDEGMWVSFWYIPVRQVVSFKKEVYRAMEFASVLFDGAK

## ORIGIN

ENLSSKSNESDKTHHTTKSTFLTKSHKPHKSRG"

Query Match 28.0%; Score 814.8; DB 1; Length 10719;

Best Local Similarity 68.7%; Pred. No. 5.2e-241;

Matches 1161; Conservative 0; Mismatches 507; Indels 22; Gaps 2;

QY 4 AAGGTTGTTATGCTCAGGGTTATCTGACAGTTTCCCAATATTGACCCCGATTTGTTTC 63  
 DB 9004 AAGGAATTTTATGCAATGCAATTAATTTTCAATTTCCCAATTTGATCCGTCATTTTGA 8945  
 QY 64 GATCGCCCTCTAGCCGTCGCTGTAATGCTGATGATTTGGTGGTTTCTTTTTC 123  
 DB 8944 AATCGGCTATCGGTTACGTTGATGCTTAATGATCTGCTGGGTTTCTTTTC 8885  
 QY 124 TATGTGTTGGCCATGCGGACCGATGCGCGGGGAGTGTGGAACGCTGACAGT 183  
 DB 8884 CCGTTGCTTGGCGGTAAGTCTATCAAGGACGAGTGGCTGACACAGACCAAGT 8825  
 QY 184 CTCGACTTGTATTTGCGCGGCTTTTATGAGTATGATGCGTGGCCGAGTTGATATG 243  
 DB 8824 GGAATGCTGTATTTTACGCTTTTATGCGCTTTTATGATGAGCTATTTGTTACGT 8765  
 QY 244 GATCTTCTACAAATTTGATCTGTTCTTCTGCTGACCTCTTATTTATTTCAAGTGTGAC 303  
 DB 8764 GTTGTTTATCAATTTGATTTATTTCTTAACAAGATCGGCTATTTATTCGTTTGGGA 8705  
 QY 304 TGGCGGATGCTCTTCCACGCGGCTTATTTGGTGTGATCAAGCCATTTCTGATATG 363  
 DB 8704 AGCGGATGCTCTTCCATGCTGCTTAATTTGATGCTATTTGATGCTATTTATTCACAGC 8645  
 QY 364 GCGTAAACCAACGACCTTCTTTTGGTGTGCGATTTGTTGGCCCTTTAGTGCAT 423  
 DB 8644 TAAATATCAAAAACGGGCTTTCTGCAACGCGGATTTTGTGCGCGCTTATTTCAAT 8585  
 QY 424 CCGTTTGGGATGGAAGATATGCTTAATTTGAATGATGAACCTTTGGGAGACGATGAC 483  
 DB 8584 TGGCTTAGGATGGAAGATGCGCAACCTTTATCAATGAGAAATGTTGGGGGCTGAC 8525  
 QY 484 GATGTCCTTGGGCTTTTATTTCTTAATGCTGACCTGCGGCTATCTTTCACA 543  
 DB 8524 GATGTCCTTGGGCTTTTATTTCTTAATGCTGACCTGCGGCTATCTTTCACA 8465  
 QY 544 GCTTTATGATTTGCTTGAAGCGGCTTCTGCTTCTTATTTCTTAATTTGTTATG 603  
 DB 8464 GCTATGCAAGCGGATTTGAAGGATGCTATTTCTTATTTCTGATTTGATATCA 8405  
 QY 604 TAAACCTGCTCCGCTAGGACGCTATCCGACGCTTTTATGCTGATACGCTATCCG 663  
 DB 8404 AAGGCTGACCAATGCGGCAACCGCTGTTATTTCTTACGCTTACGATTTTCCG 8345  
 QY 664 CTTCCTTGTGAATAGCTCCGTAGACCAAGTCTCACTGGTCT--GTTGGTGGCTT 720  
 DB 8344 TTTTATGCTGATTTTTCGTAGGCTGATGCCCAATTTGCTTATTTTGTGACACA 8285  
 QY 721 CATTTGAATGGGCAATCTCTCTTACTATGATGATGATGATTTTGAATGATG 780  
 DB 8284 TATCTCATGGAACAATTTCTTCCACACCAATGATTTCTATTTGGTGGCTATCAATG 8225  
 QY 781 TTGCTTTCAAGCGGCTTTGATCAAGACGCTGATGACGAAATAGGATGATAGT 840  
 DB 8224 AGTGGATATCAAAAGT-----CGGTAATAAAAGGAATATTTAT 8184  
 QY 841 GAAACGATTTATGATCTTTTGTACGCGCATGCTGATCAAGTGTGTTGTTGATAAATGA 900  
 DB 8183 GAAACGATTTATGAACTTTTGTACCAAGTATGCTGATGATGATGTTGGTGTGATAAATG 8124  
 QY 901 AGGAGGAGGAGGATGTTGATGCTGATGATGATGATGATGATGATGATGATGATGATG 960  
 DB 8123 AGGAGGAGGAGGATGTTGATGCTGATGATGATGATGATGATGATGATGATGATGATG 8064  
 QY 961 CAATCAGTTTCTCTAGTACTACAGCAAGATTTTGGAAAGCTCCCTAGCCGAGTT 1020

DB 8063 TAAACAATTTCCGTTGATCAACACAGCAAAAGCTATTGAAAACCGGCAATTCAGAAAT 8004  
 QY 1021 GCTGGCTATATTCGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
 DB 8003 TTTAGGTTATATTCGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 7944  
 QY 1081 CTGGATGCTATATGCAATTTTAAACCAAGATGCTCAACATCTTACCTGTAAGATGA 1140  
 DB 7943 CTGGATGCTATATGCAATTTTAAACCAAGATGCTCAACATCTTACCTGTAAGATGA 7884  
 QY 1141 GATGACATGGAAGCGGCTGATGCTTCAAGGCTAGACCTTGGGCTTAACTGATGATG 1200  
 DB 7883 CGATGATATGGAAGCGGCTTATGCTTCAAGGCTGATGCTGATGCTGATGCTGATGCT 7824  
 QY 1201 TCAATATGACGCTGTAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260  
 DB 7823 AACCGTCAATTCGCAAAATCTCAATATTCGCGAACGCAATGATGATGATGATGATGAT 7764  
 QY 1261 TGAATTTCTTAACTTCAATTCGCGTGAATTTCAATGAGGCTGTTTGGCCTTGCAT 1320  
 DB 7763 TGAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7704  
 QY 1321 GTAGACCATCATTTTCAATGCTGAGGAGATTCCTTATTCACAGTATCAAGCTGATG 1380  
 DB 7703 GCACACCAACACCTTCTCAATCTGCTGATACCTCTATCTCAAGATATCAAGCTGATG 7644  
 QY 1381 ATGATATGCTGCTTGGGCTTGAATTTCAACATGCTGAGGCTTATGTTCTTCTGCT 1440  
 DB 7643 TGTGATGCTGCTTGGGCTTGAATTTCAACATGCTGAGGCTTATGTTCTTCTGCT 7584  
 QY 1441 GATGACATGATCAACAGGAAAAAGCCGGCTTGGCTATCAACAGTCTGATGATGATGAT 1500  
 DB 7583 GATGACATGATCAACAGGAAAAAGCCGGCTTGGCTATCAACAGTCTGATGATGATGATGAT 7524  
 QY 1501 CATTTACCAAGATCAATGCAATTTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 1560  
 DB 7523 CATTTATGAAGATCAATGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 7464  
 QY 1561 AGGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620  
 DB 7463 GTTGCACCAATTAACATTTTACCTGATGATGATGATGATGATGATGATGATGATGAT 7404  
 QY 1621 CACTTGAATGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680  
 DB 7403 GACTATGATGATTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7344  
 QY 1681 TTCAGTCTAA 1690  
 DB 7343 CTCTGTGTA 7334

RESULT 14  
 AF064791 1893 bp DNA linear BCT 31-DEC-2001  
 LOCUS AF064791  
 DEFINITION Pasteurella multocida thymidylate synthase (thy) gene, complete cds.  
 ACCESSION AF064791  
 VERSION AF064791.1 GI:18000047  
 KEYWORDS  
 SOURCE Pasteurella multocida  
 ORGANISM Pasteurella multocida  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales; Pasteurellaceae; Pasteurella.  
 PASTEUR  
 1 (bases 1 to 1893)  
 MORENO, J.A., BOSCH, M., BADIOLA, I., LLAGOSTERA, M. and BARBE, J.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-MAY-1998) Departamento de Genética y Microbiología (Unidad de Microbiología), Universidad Autónoma de Barcelona, Edificio Cn, Bellaterra, Barcelona 08193, Spain  
 FEATURES  
 source  
 1..1893  
 /organism="Pasteurella multocida"  
 /mol\_type="genomic DNA"

```

gene      /strain="PM25"
          /specific_host="bovine"
          /db_xref="taxon:747"
826. .1677
          /gene="hlyA"
CDS       826. .1677

```

## ORIGIN

Query Match	26.6%	Score 775.2;	DB 1;	Length 1893;
Best Local Similarity	67.9%	Pred. No. 7.8e-229;		
Matches 119;	Conservative	2;	Mismatches 505;	Indels 22;
				Gaps 2

QY	46	TGACCCCGATATGTTTTCGATCGGCGCCCTGACGGGTGGCGTGGATGATGATAT	105
Db	49	TGATCCCGATTTTTGAAATCGGCTCTATCGGTTTACGTGGTATGGCTTATGTA	108
QY	106	GGTGGGTTTCCTTTTGGCTATGTGGTTGGCAATGCGGAGCGATCGCGGGAGTGG	165
Db	109	GCTGGGGTTTCTCTTGGCCCGTGGCTTCGGGTGAAGATGTCTAACCAACAGGCA	168
QY	166	TTGACGGGATGAGCAAGTCTCGATCTGTATATGCGCGGCTTTTAGGATGATACGG	225
Db	169	CTGGACAACAGACCAAGTGAATGCTGTATTTATTAACGGTTTATAGGGCGTTT	228
QY	226	TGCGCGATGGTATGTATGTATCTTCACAAATTTTGAATCTGTTCTGTGACCTCT	285
Db	229	TGGCGCATTTGGTATCGTTGTGTTTATCAATTCATATTTCTTACAAAGATCTG	288
QY	286	TTTATTTCAATGTCGACTGCGGAGATGCTTCACACGGCGGCTATTTGGGTGAT	345
Db	289	TTTATTCGTTTGGGAAGGGGATGCTTTTCATGCGGCTTAAATTTGGTATGTA	348
QY	346	CGCCATGTTCTGTATGCGCGTAAACCAACGACACTTCTTTGGTGTGCGGATTT	405
Db	349	AGCATTGTATTCACGGCTTAAATACAAAAACGGGGTTTCTGGCAACGGCGGAT	408
QY	406	TGCCCCTTTAGTCGATTCGGTTTGGGATGGGAGCTATCGGTAACTTTATGAA	465
Db	409	GGCACCGTGTATCCATTTGGCTTAAGTATGGGACGGATCGGCACTTTATCAAG	468
QY	466	ACTTTGGGAGAGATTAACGATGTCCTTGCGCTTTTATCTCCCTAATGAGGCCA	525
Db	469	ATTGTGGGGGGGTGTACAGATGTGCCATGGCACTCTTATCCATCAGGGGATAT	528
QY	526	GCGCGCCATCTCTCAAGCTTATGAATGCGCTTAAAGCGGTGTTCTGTCTTAT	585
Db	529	ACCAACGGCATCMTCAAGCTAATAGACCGGTATTAAGAAAGGATGTGTCAT	588
QY	586	TCTTAATATGTTTATTTGGTAAACCTCGATCGGTAGACAGCTATCGGACTGTT	645
Db	589	CTCGAATTTGTATATCAAAAACCTCGAACCAATCGGTGCAACSTTGGTTTAT	648
QY	646	TGATATACGATCATTTCCGCTTCTTTGGAAATACGTCCGTGAGCCAGATGCT	705
Db	649	CGGTTACGGTATTTTCGTTTATCGTTGAGTTTTCGTGAGCTGATGCCAATTA	708
QY	706	TCT---GTTTGTGCTTCAATTTCAATGGGGAAATCCTCTCTTACCTATGAT	762
Db	709	CTTATATTTTGTCAACATCTCATATGGGACAAATTTCTTCAACCAATGATTT	768

QY	763	CGGATATTTGAGNATGCGTTGGTCTTACCAACGGCGGTTTGATCAAGACCGTGTAGCAGC	822
Db	769	TGCTGCGGTAATCATGTTAGTGCGATATCAAAAGTG-----CGGGTA	809
QY	823	AAAAATAGGGTAGTTAGGTGAAACAGTATTTTATGATCTCTTGTGACGCGATCGTGCATCAAG	882
Db	810	AAAAAAGTAATTTTATGAAACACTATTTTGAACCTTTGCCAACGTATCGTTGATGAAG	869
QY	883	TGTTTGGGTTGAAAATGAAACGACGCGACAGCGTTGTTGACTGTGATTAATGCCGATTT	942
Db	870	TGTTTGGGTTGAAAATGAAACGACCTGAAAAAGTTGTTGACCGGATGTGATGCAGACT	929
QY	943	GACCTACGATGCGGCAACATTCAGTTTCCATGATGACTACGACGAAGAGTTTGGAA	1002
Db	930	CACCTACGATGTTGCGATTAACCAATTTCCGTTGATCAACACGCAAAAGCTATTGGAA	989
QY	1003	AGCTGCCGTAGCCGAGTTCCTCGGCTATATTTGTTGTTACGATATATGCGGCGGATTTTCG	1062
Db	990	AGCGGCATTTGCAGAAATTTTATAGTTATATCCGTGGCTATATGACATATGCGGAGATTTTCG	1049
QY	1063	CCATTTAGGTATCCAAACCTTGGATGCTATATGCCAATTTAAACCAAGATGGCTCAACA	1122
Db	1050	CAAACTCGGCAACAAACCTGGGATGCGCAATGCTAATGAAAAACAGTATGGCTCAATTA	1105
QY	1123	TCCTTACCGTAAAGGTGAGATGATCATGGGACGGGTGATGTTGTCAGGGTAGAGCTTG	1182
Db	1110	TCGCGACCGCAAGGACCGAGATGATGGACGTTGTTATATGGCTTACAAGGCGCTGATG	1166
QY	1183	GGCTTAGCCTATGCTGCTCATATTTGACACAGTTGAAAAAGATTTGTTGATTTTGAGCCG	1242
Db	1170	GCGCAACCGAACGCTGAAACCGTGCATCATTTGCGCAAAATCGTCATTAATTCGCGCA	1222
QY	1243	TGGCGTTGATGACCGAGGTGAAATTTCTTAATCTTCACAATCCGGTGAATTTACATGGG	1302
Db	1230	CGGCAATTGATACCGTGGTGAATCATGACCTTCTTCAACCCCGTGAATTTGAACTGGG	1289
QY	1303	GTGTTTGGGCCCTGCAATGTCACAGCAATTTTTCATGTGCGGGGATACCTGTATCT	1362
Db	1290	TTGCTTACGTCTTGTGATGACAACCACACTTCTCATTTACTCGGTGATACCTCTATCT	1349
QY	1363	CAACAGTACTACGGCTTCATGTGATGTGCCCTTGGGGTTGAAATTTCAACATGTGTACAGT	1422
Db	1350	CACACGTTATCAGCGTCTTGTGATGTGCGGCTTGGCTGAATTTCAACAAATTCAGT	1409
QY	1423	TTATGTGTTCTTGCGCTGATGCGACAGATCAACGGGAAAAAGCCGGGCTTGGCGTATCA	1482
Db	1410	GTTCACTCTTGTGCTTGTGATGCGACAAATTAACGGGCAAAAAAGGGGTAAAGGGTATCA	1463
QY	1483	CAAGATCGTCATGTGCGCATTTTACCAAGATCAACTGGAATTTGATGCGCGATGTGCACCT	1542
Db	1470	CAAAATCATCATGACACATTTTATGAAGATCACTGTATCTCATGAAAAATGTGCAAAT	1522
QY	1543	AAAAAGTAGCATTTCCAGCGGCTCAGTTCCATATCAATCCAAAGATTAATAACACTGCA	1602
Db	1530	AAAAAGTAGACCTTCCCGTTGGCCACATTTGCAATTAACCCCGAGATTAATAACCTCGA	1588
QY	1603	GGATTGTGAAACTTGGGTCACCTTGTGATGATTTTGAAGTCAACCGGATATCACTTCAACGA	1662
Db	1590	AGATTTAGAAACTTGGGTGACTATGATGATTTTAAAGTCAACGGGCTATCATGACGACGA	1644
QY	1663	TCCTATTTCAATACCGTTTTCAGTCTAA	1690
Db	1650	TGCCATTTAAATACCTTCTCTGTGTAA	1677

RESULT 15	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE
U32772	U32772	Haemophilus influenzae Rd section 87 of 163 of the complete genome.	U32772.1	GI:1573918		Haemophilus influenzae Rd Kw20

ORGANISM	Haemophilus influenzae Rd KW20
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales; Pasteurellaceae; Haemophilus.
AUTHORS	1 (bases 1 to 11545) Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirtnes, B.F., Kerlavage, A.R., Bult, C.J., Tomb, J., Dougherty, B.A., Merrick, J.M., McKenney, K., Sutton, G.G., Fitzhugh, W., Fields, C.A., Gocayne, J.D., Scott, J.D., Shiley, R., Liu, L.I., Glodek, A., Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E., Cotton, M.D., Utterback, T., Hanna, M.C., Nguyen, D.T., Saudek, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhmann, J.L., Geoghagen, N.S., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M., Smith, H.O. and Venter, J.C.
TITLE	Whole-genome random sequencing and assembly of Haemophilus influenzae Rd
JOURNAL	Science 269 (5223), 496-512 (1995)
MEDLINE	95350630
PUBMED	7542800
REFERENCE	2 (bases 1 to 11545) Tatusov, R.L., Mushegian, A.R., Bork, P., Brown, N.P., Hayes, W.S., Bordovsky, M., Rudd, K.E. and Koonin, E.V.
AUTHORS	Metabolism and evolution of Haemophilus influenzae deduced from a whole-genome comparison with Escherichia coli
TITLE	Curr. Biol. 6 (3), 279-291 (1996)
JOURNAL	96398784
MEDLINE	8805245
REFERENCE	3 (bases 1 to 11545) White, O., Clayton, R.A., Kerlavage, A.R. and Fleischmann, R.D.
AUTHORS	Direct Submission
TITLE	Submitted (25-JUL-1995) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
JOURNAL	4 (bases 1 to 11545) White, O., Clayton, R.A., Kerlavage, A.R. and Fleischmann, R.D.
AUTHORS	Direct Submission
TITLE	Submitted (27-SEP-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
JOURNAL	The H. influenzae sequence has been updated by R. Fleischmann. New database matches have been assigned, product names have been improved, and a number of frame shifts have been corrected. We gratefully acknowledge the work of Tatusov et. al. We have incorporated their annotation into the /notes fields of the corresponding H. influenzae genes
REMARK	5 (bases 1 to 11545) White, O., Clayton, R.A., Kerlavage, A.R., Fleischmann, R.D., Peterson, J., Hickey, E., Dodson, R. and Gwinn, M.
COMMENT	Direct Submission
FEATURES	Submitted (28-MAY-1998) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA The whole genome was shifted by 588 nucleotides for a new start On Sep 30, 1996 this sequence version replaced gi:1221622. Location/Qualifiers 1..11545 /organism="Haemophilus influenzae Rd KW20" /mol_type="genomic DNA" /db_xref="taxon:71421" complement(122..604) /gene="HI0899" complement(122..604) /gene="HI0899" /note="similar to GB:D10483 SP:P00379 GB:J01609 GB:V00276 GB:X05108 percent identity: 53.16; identified by sequence similarity; putative" similarity; putative" /codon_start=1 /transl_table=11 /product="dihydrofolate reductase (folA)" /protein_id="AAC22559.1" /db_xref="GI:1573919" /translation="MTFSLIVATLTANNVYGNQNPWHLPAFLAFRQNTTGPVINGRTFESIGRLPKRTNIVTSRLPFEEGVWDSFESAVNPFVDPDELWIGGELPFKQYLPKADKLYLTQITELDGTFFQLNMEWEIEIDFYEKADQNRDYCRPLILTRK"
gene	705..1811 /gene="HI0900"
CDS	705..1811 /note="similar to GB:U00096 SP:P07005 PID:1208988 PID:1552810 PID:1786437 percent identity: 65.57; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="glutamate 5-kinase (gamma-glutamyl kinase) (proB)" /protein_id="AAC22560.1" /db_xref="GI:1573920" /translation="WNKKTIVKFGTSTLTQSPPLNSPHMMEIVRQIALQINDPFRIVISGALAAGRHYLNHPOLPPTIASKQDILLAVGQSOILOAEKFLPYDHIHQLLTRADIEDRRFLNARDTLVYALDNHIIPIVNDVAVAEIVGNDNLSALVALIVQABQVLTLDQGLPDSDRKNPEAKLIPVEQITDHSIAGSGTNGTGMWTKIIMADVATRSIGETITAPKRPVNIADLAIEQNGTGFIAHQSDRLSRKQMLFAASAGLITINDGAONALIEONKSLPVGIIIVGERSRSGVVKITQSGKQIALGMPRTNSDALQILGKRSADILENVLGYEYGAVMHRDMILLS"
gene	1892..2482 /gene="HI0901"
CDS	1892..2482 /note="similar to PID:882723 GB:U00096 SP:Q46930 PID:1789194 percent identity: 77.71; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAC22561.1" /db_xref="GI:1573921" /translation="MIDFGYRPVNGIYICNRKGQVLNAKRCGNSQWOPQGINDESAEQVREIHEHVEGLQPKDRLVLYVSGHRYKLPKRLRDSKPMGICQKQWFLQLVSDERKINNOTTSPPEDGRWYSFMYPRQVYSPFDYRKYKMEFASLITFDNPLIFSASREANSLSHSANKTSQTKTKHFASKQ"
gene	2482..3276 /gene="HI0902"
CDS	2482..3276 /note="similar to GB:U17717 SP:O57883 PID:1591145 percent identity: 28.84; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAC22562.1" /db_xref="GI:1573922" /translation="MFTFILLCLVGLAGLAGLFGIGGLVIVPTLVLYLPIDVPLSLMSTALGTSFATVITVIGISAOHRKLGNIQAARILAPVIMLVFCGLFGRIDREISAKIPACLVVYATKQWLSIKKQVTTKSLTPLSVVGILIGMASSAAGIGGGEFVPLTARGINIKOALIGSAPCGMILISGMSFVIGSGNLMREYSGLIYLPAYLITATSPFPTSKAGATKALPVSTLKKGFALFLIVVAINMFLK"
gene	3285..4091 /gene="HI0904"
CDS	3285..4091 /note="similar to SP:P37149 GB:U12289 PID:516036 PID:882720 GB:U00096 percent identity: 62.78; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="proliprotein diacylglycerol transferase (lgt)" /protein_id="AAC22563.1" /db_xref="GI:1573923" /translation="WNSYLYLLPHRPSIFTLGDSNIGLRMYGLMYLLGFYFARMLAVBRANPNSGWATVDQDSLIFNGFMGVPIGCGVGVFYNLHFLQEPYLYLFWMGGMSPFGGLIGVIMWTYSQKRNFWQCTFAPALPFGIGLGRINPILNLEMGSETNVPVAMIPNDPILLBRHSOLYEALFELGVLFTINITIKKRPASVAGELIYGVRFRIVYVREBVENFPGIITRQGLCLPMIIGAFINAMAYSRSAYIK"
gene	4101..4952 /gene="HI0905"
CDS	4101..4952 /note="similar to GB:J01710 SP:P00470 PID:147987 PID:42685

PID:882719 percent identity: 35.32; identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="chymidylate synthetase (thya)"  
 /protein\_id="AAC22564.1"  
 /db\_xref="GI:1573924"  
 /translation="MKQYLELCRIIVSEGEVWNERGKHCITVINADLBYDVANNQFPLITRKSYKRAIAIEFLYIRGYDNADAPALGTPWDANANENAAWLANPBRGVDMDGRVYGVQAMRKNGETIDQLKIVNNLTGKIDIDGELITFPNPEPDLGCLRPCMHTTFESIVLETHLITSYORSODVPLGKINPNOIOVFTPLAIAOITGKKAKAYKIVNAHIEDQLELMRDVQLKRBPPLPKLEINPDIKLELDELTWVTMDPRKVVGYQSHERP I K Y P S V"  
 /gene=".5473  
 /gene="HI0906"  
 /note="similar to Sp:P30134 PID:296183 PID:987637  
 PID:103148 GB:U00096 percent identity: 57.41; identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAC22565.1"  
 /db\_xref="GI:1573925"  
 /translation="MDAAKVRSEPDENRMYALBLADKALGLIPGAVVVDARNTIGENWNLISVSDPTAAHEITLRNGANININRYRLNLTGTLVLTPECTMGALIHRIKRLVFGASDYKTAIGSRFHPFDYKNNHITLTVTSVLAESCQKSLTFPOKREKKIERKALKSLSDK"  
 complement(5501..5815)  
 /gene="HI0907"  
 complement(5501..5815)  
 /gene="HI0907"  
 /note="hypothetical protein; identified by Genemark; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="H. influenzae predicted coding region HI0907"  
 /protein\_id="AAC22569.1"  
 /db\_xref="GI:1573930"  
 /translation="MENKARYOKAVNITVDLQSPAKIIRKGLAINEINOKENRIPQEFHGRFRIGMNTNLSIFETANAIVQGLIFRQTELLKLIQEFEPVOTPEITINP GP"  
 5895..6197  
 /gene="HI0908"  
 5895..6197  
 /gene="HI0908"  
 /note="hypothetical protein; identified by Genemark;  
 Query Match 25.6%; Score 744.2; DB 1; Length 11545;  
 Best Local Similarity 66.7%; Pred. No. 4.9e-219;  
 Matches 1134; Conservative 0; Mismatches 538; Indels 27; Gaps 4;

gene  
 CDS

gene  
 CDS

gene  
 CDS

QY 296 GTGTGACTGGCGGCGATGCTCTTCCAGCGCGCTTATTGGGTGATGACCGCATGTTTC 355  
 DB 3573 GTTTGGGAAGGTGGAATGTCTTCCAGCGTGGCTTAAATGGTGAATGTTGGTCATGAT 3632  
 QY 356 TGGTATGGGCGTAAACCAACGACACCTCTTGTGGTGGCGAATTTGTTGCCCTTGA 415  
 DB 3693 TGGACATTTATCTCAAAAACGTAATTTTGGCAAAACGCGTGAATTTGTGGCGCTTTG 3692  
 QY 416 GTGCCATTGCGTTTGGGATGGAGCGTATCCGTAACTTTATGAATAGTAACTTTGGGGA 475  
 DB 3693 ATTCCGTTTGGTTTAAAGTTTAAAGCCCTTTTAAAGGCGCTGGTGTGTTACGATTTGAT 3752  
 QY 476 CGAGTAACGATGTGCGCTTGGCGCTTTGTATCCCTAATGTTGCGC---CACTGCGCGC 532  
 DB 3753 CGCGAAACGAATGTGCGCTTGGGCAATGATTTCCCGAATGATCTCTTTTACCTGCTCGT 3812  
 QY 533 CATCCTTCACGCTTTATGATTTGGCCTTAAAGGCGGTGCTGTTCTTATCTTAT 592  
 DB 3813 CATCATCAACAATTATGAAGCCCTTTTAAAGGCGCTGGTGTGTTACGATTTGAT 3872  
 QY 593 TGGTTATTTGTTAAACCTGTCGCGCTAGGCAAGCGTATCCGACTGTTTATAGTGATAC 652  
 DB 3873 ATTTTATTAATAAAACACGTCGAATGCGCTTCTGTGAGGTTTATCTTAAATGGTTAT 3932  
 QY 653 GGTACATTCCGCTTCCTTGTGGAATAGTCCGTGAGCCAGATGCTCAAGTTGGTCTGTT 712  
 DB 3993 GCGCGTCCCTGTTTATTTGGAATATGTGCGTAACCTGAAG-----TTGAAATTTTC 3986  
 QY 713 GGTGCTTCATTTCAATGGGCAAAATCCTCTTACCTAATGATGATCATCGGTATTTG 772  
 DB 3987 TTTGGGATTTATTAACGAGGCAAGCCCTTTGCTTGCCGATGATTAATGGTGTGCTTTC 4046  
 QY 773 ATGATGTTTGGTCTTACGAGCGCGTGTGTATCAAGACCGGTAGACGAAATAGGGT 832  
 DB 4047 ATTAATGCTTGGGCTTATTAACGC-----AAAGTGGGCTCATAAATAGAG 4094  
 QY 833 AGTTAGTGAACAGATATTAGATCTTGTGAGGCGCATGTCGATCAAGGTGTTGGGT 892  
 DB 4095 ATTTTATGAAGCAATATCTTGACCTTGTGCGCGCATGTTGATGAAAGGGGAATGGGTT 4154  
 QY 893 GAAATGAGACGAAGGGGCAACGTTGTTGACTGTGATTAATGCGAATTTGACCAAGAT 952  
 DB 4155 GCTATGACGTAACGAGTAACGATTTGCTCAAGGTCATTAATGAGATTTGAAATATGAT 4214  
 QY 953 GTGGCAACAATCAAGTTCTCTAGTGACTACAGCAAGATTTTGGAAAGCTGCCGTA 1012  
 DB 4215 GTGGGAATATCAATTTCCGCTGATTAATCAAGGTAATGTTGGAAGCGGCGATT 4274  
 QY 1013 GCCGATGCTCGGCTATATTCGTGTGTTACGATTAATGCGGCGATTTTCCCAATTAGT 1072  
 DB 4275 GCTGAATTTTAAAGTTATATTCGTGATATGACATATGCGCTGATTTCCGCGCACTTGGC 4334  
 QY 1073 ACCAAACCTGGGATGCTAATGCAATTTAAACCAAGATGGCTCAACAATCCTTACCGT 1132  
 DB 4335 ACGAAACTTGGAGTCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 4394  
 QY 4395 AGAGCGTGTATGATATGAGGCGGCGTATATGTGTGCAAGGCAAGCATGCGTAAGCT 4454  
 QY 1193 GATGTGTCAATTTAGCAAGTGAATAAAGTGTGATGATTTGAGCGCGTGGTAT 1252  
 DB 4455 AATGAGAAACTATTCGATCACTACGTGTAATAATTTGTATTAATCAAGAAAGATATGAT 4514  
 QY 1253 GACGAGGTGAATTTCTAATCTTCAACAATCGGSGTGAATTTCAATGSGGTTTGGCG 1312  
 DB 4515 GATAGAGAGATTTTAACTTTTAACTTCCGAGGAATTTGATCTTGGTGTCTTGT 4574  
 QY 1313 CTTTGAATGACACCATCATTTTCAATGCTGGGAGATACCTTGTATCTCAAGATGACT 1372  
 DB 4575 CTTGTATGATAGCAATCACTTTTCTGTGTGGCGATACCTTAACTTAACTTAAGCAT 4634  
 QY 1373 CAGCGTTCATGATATGTCCTTGGGCTTGAATTTCAATGATGTGACAGTTATGTGTTTC 1432

```
Db 4635 CAGCGTTCCCTGATGATGATTCGCTTGATGAATTCATCAATCAGGTTTACCTTC 4694
Qy 1433 CTGGGCTGATGGCAGATCAAGGAAAAAGCCGGCTTGCGTATCAAGATCGTC 1492
Db 4695 TTAGCACTTATGGCTCAGATCAGGCAAAAAAGCCGCAAGCATATCATAAATTGTG 4754
Qy 1493 AATGCCACATTTACCAAGATCAACTGCAATGATGCGCGATGCGACTAAACGTGAG 1552
Db 4755 AATGCCATATTTATGAGATCAGCTTGAGTTAATGCGTGACGTACACTTAAACGCGAG 4814
Qy 1553 CCATTCCCAAGCGCTCAGTTCCATTCATTCATAAGATTAAACACTGCAAGATTGGAA 1612
Db 4815 CCTTCCCATTTACCAAAATTAGAAATTATCCAGATATATAAAACGCTGAAGATTGGAA 4874
Qy 1613 ACTTGGGTCACTTTGATGATTTTGAAGTCAACCGGATATCAGTTCACGATCCTATTCAA 1672
Db 4875 ACTTGGGTCAAGATGATGATTTTAAAGTGTGGCTATCATCCACGAAACCAATTAAA 4934
Qy 1673 TACCGTTTTCAGTCTAAT 1691
Db 4935 TATCTTTTTCGCTTAAT 4953
```

Search completed: July 31, 2004, 15:38:59  
Job time : 7462 secs

This Page Blank (usp:0)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2004, 07:32:25 ; Search time 761 Seconds  
(without alignments)  
16239.170 Million cell updates/sec

Title: US-09-700-712a-1  
Perfect score: 2909  
Sequence: 1 gaggaagcttgcctgcctc.....cagcgaatcggcgctgcag 2909

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Listing first 45 summaries

Database : N\_Geneseq\_290and04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2003as:\*  
8: geneseqn2003bs:\*  
9: geneseqn2003cs:\*  
10: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2909	100.0	2909	3	AAZ40646
2	1222	42.0	1222	3	AAZ40646
3	1112.6	38.2	1149	7	ACA53067
4	842.4	29.0	852	2	AAZ24801
5	842.4	29.0	852	7	ACA53066
6	838	28.8	838	3	AAZ40647
7	744.2	25.6	110000	2	AAT42063_09
8	505.2	17.4	852	7	ACA42822
9	462.4	15.9	852	7	ACA44455
10	403.8	13.9	849	7	ACA44455
11	292.2	10.0	807	7	AAZ37118
12	278.2	9.6	110000	7	ACF67367_27
13	278.2	9.6	110000	7	ACF65386_0
14	278.2	9.6	110000	7	ACF65386_1
15	277.8	9.5	876	7	ACF69784
16	248.2	8.5	876	5	AAH81383
17	208.6	7.2	834	6	ADA32523
18	208.4	7.2	834	6	ABO90157
19	161.2	5.5	852	3	AAZ53292
20	156.4	5.4	852	3	AAZ53291
21	156.4	5.4	102634	3	AAAB1464
22	156.4	5.4	110000	3	AAAB1490_10
23	156.4	5.4	349980	3	AAZ21609

24	152.2	5.2	110000	6	ABA92787_4	Continuation (5 of
25	149.2	5.1	849	7	ABZ40310	Abz40310 N. gonorr
26	149.2	5.1	852	3	AAZ53290	AAZ53290 Neisseria
27	148.6	5.1	1515	4	AAZ61079	AAZ61079 P. putida
28	145.4	5.0	420	6	ABN26154	ABN26154 Human ORF
29	142.2	4.9	96109	4	AAZ28548	AAZ28548 Genomic E
30	137	4.7	420	6	ABN26172	ABN26172 Human ORF
31	84	2.9	2440	9	ABT41887	ABT41887 Toxicity
32	84	2.9	2440	9	ADBS8174	ADBS8174 Toxicity-
33	84	2.9	2440	9	ADBS2673	ADBS2673 Primary r
34	83.6	2.9	4473	5	AAZ88385	AAZ88385 DNA encod
35	79	2.7	2075	9	ADBS8991	ADBS8991 Toxicity-
36	79	2.7	2075	9	ADBS3745	ADBS3745 Primary r
37	77.8	2.7	3950	7	ABT42286	ABT42286 Toxicity
38	75.4	2.6	2208	4	ABA09555	ABA09555 Human Na-
39	75.4	2.6	2208	4	AAKS2748	AAKS2748 Human pol
40	75.4	2.6	2280	4	ADBS0585	ADBS0585 Ovarian C
41	75.4	2.6	2288	2	AAV59498	AAV59498 Human sod
42	75.4	2.6	2380	4	AAKS1764	AAKS1764 Human pol
43	75.4	2.6	2520	4	AAK94823	AAK94823 Human ful
44	75.4	2.6	3952	8	ACH03831	ACH03831 Human CDV
45	75.4	2.6	4137	3	AAZ75484	AAZ75484 DNA encod

## ALIGNMENTS

RESULT 1	AAZ40646	standard; DNA; 2909 BP.
ID	AAZ40646	
XX	AAZ40646;	
AC		
XX		
DT	08-MAR-2000	(first entry)
XX		
DE		Nucleotide sequence of V. cholerae thya gene.
XX		
KM		Vibrio cholerae; thya-negative strain; Deltathya; thya gene; vaccine;
XX		cholera; antimicrobial therapy; ss.
OS		Vibrio cholerae.
XX		
PN	MO9961634-A1.	
XX		
PD	02-DEC-1999.	
XX		
PF	21-MAY-1999;	99WO-EP003509.
XX		
PR	26-MAY-1998;	98SB-00001852.
XX		
PA	(SBLV-) SBL VACCIN AB.	
XX		
PI	Carlin N, Lebens MR;	
XX		
PS	WPI; 2000-062719/05.	
XX		
PT	P-PSDB; AAY59126.	
XX		
XX	New Vibrio cholerae strain defective in the thya gene, for use in	
XX	vaccines and for recombinant protein production.	
XX	Claim 9; Fig 1; 42pp; English.	
CC	The invention provides a method for producing a Vibrio cholerae thya-	
CC	negative strain which is Deltathya strain lacking thya gene functions.	
CC	The method comprises site-directed mutagenesis of the V. cholerae	
CC	chromosome to delete and/or insert nucleotides at the thya locus; The V.	
CC	cholerae thya-negative strains are used: for overproduction of	
CC	recombinant proteins; and in vaccines to prevent or treat cholera (or	
CC	other diseases if engineered to express the appropriate proteins). The	
CC	thya gene is also useful for insertion of foreign genes, in a selective	
CC	and site-specific manner, and the proteins expressed by the thya gene or	
CC	by its 5'-flanking region, are useful in research and as targets for	
CC	antimicrobial therapy. When used for recombinant protein production, V.	

CC cholerae provides high yields with secretion of products into the culture  
 CC medium for ease of subsequent recovery. The thA-negative strain can be  
 CC maintained by thymine complementation, eliminating the need for  
 CC antibiotic selection. The present sequence represents the nucleotide  
 CC sequence of the V. cholerae thA gene  
 CC  
 XX

SQ Sequence 2909 BP; 734 A; 644 C; 737 G; 794 T; 0 U; 0 Other;

Query Match 100.0%; Score 2909; DB 3; Length 2909;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 2909; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGAAAGTTTGTATGCTCAGAGGTTATCTGAGATTTCCCAATATTTGACCCCGATTTGT 60  
 DB 1 GAGAAAGTTTGTATGCTCAGAGGTTATCTGAGATTTCCCAATATTTGACCCCGATTTGT 60

QY 61 TTGATTCGGCCCTTAGCGGTCGCTGGTATGGCTTATGATATTTTGGTGGTTCTCTTT 120  
 DB 61 TTGATTCGGCCCTTAGCGGTCGCTGGTATGGCTTATGATATTTTGGTGGTTCTCTTT 120

QY 61 TTGATTCGGCCCTTAGCGGTCGCTGGTATGGCTTATGATATTTTGGTGGTTCTCTTT 120  
 DB 61 TTGATTCGGCCCTTAGCGGTCGCTGGTATGGCTTATGATATTTTGGTGGTTCTCTTT 120

QY 121 TGCTATGTGTGGCCATTCGCGAGCGGATGCGCGGCGAGTGGTGGAGCGGAGCA 180  
 DB 121 TGCTATGTGTGGCCATTCGCGAGCGGATGCGCGGCGAGTGGTGGAGCGGAGCA 180

QY 121 TGCTATGTGTGGCCATTCGCGAGCGGATGCGCGGCGAGTGGTGGAGCGGAGCA 180  
 DB 121 TGCTATGTGTGGCCATTCGCGAGCGGATGCGCGGCGAGTGGTGGAGCGGAGCA 180

QY 181 AGTCTCTGACTGTATTCGCGGCTTTTAAAGTATGATCGGTGGCCGAGTGGTGA 240  
 DB 181 AGTCTCTGACTGTATTCGCGGCTTTTAAAGTATGATCGGTGGCCGAGTGGTGA 240

QY 241 TGTGATCTTTCACAAATTTTGAATCTGTTCTTGTCTGACCTCTTTATTTATCAAGTGTG 300  
 DB 241 TGTGATCTTTCACAAATTTTGAATCTGTTCTTGTCTGACCTCTTTATTTATCAAGTGTG 300

QY 241 TGTGATCTTTCACAAATTTTGAATCTGTTCTTGTCTGACCTCTTTATTTATCAAGTGTG 300  
 DB 241 TGTGATCTTTCACAAATTTTGAATCTGTTCTTGTCTGACCTCTTTATTTATCAAGTGTG 300

QY 301 GACTGCGGCGGATGCTCTTCACGCGGCGCTTATGGGTGATCACGCGCATGTCGGTA 360  
 DB 301 GACTGCGGCGGATGCTCTTCACGCGGCGCTTATGGGTGATCACGCGCATGTCGGTA 360

QY 301 GACTGCGGCGGATGCTCTTCACGCGGCGCTTATGGGTGATCACGCGCATGTCGGTA 360  
 DB 301 GACTGCGGCGGATGCTCTTCACGCGGCGCTTATGGGTGATCACGCGCATGTCGGTA 360

QY 361 TCGCGGTAAACCAACGCACTTCTTGTGTGGTGGCCGATTTTGTGCCCCCTTTAGTGC 420  
 DB 361 TCGCGGTAAACCAACGCACTTCTTGTGTGGTGGCCGATTTTGTGCCCCCTTTAGTGC 420

QY 361 TCGCGGTAAACCAACGCACTTCTTGTGTGGTGGCCGATTTTGTGCCCCCTTTAGTGC 420  
 DB 361 TCGCGGTAAACCAACGCACTTCTTGTGTGGTGGCCGATTTTGTGCCCCCTTTAGTGC 420

QY 421 ATTGCGTTTGGGATGCGAGCTATGCGTATGCTTATGAAATGATGAACTTTGGGGACGAGT 480  
 DB 421 ATTGCGTTTGGGATGCGAGCTATGCGTATGCTTATGAAATGATGAACTTTGGGGACGAGT 480

QY 421 ATTGCGTTTGGGATGCGAGCTATGCGTATGCTTATGAAATGATGAACTTTGGGGACGAGT 480  
 DB 421 ATTGCGTTTGGGATGCGAGCTATGCGTATGCTTATGAAATGATGAACTTTGGGGACGAGT 480

QY 481 AAGCGATGCGCTTGGGCTTTGTATTCCTTAATGATGCGGCGGCGCATCTTC 540  
 DB 481 AAGCGATGCGCTTGGGCTTTGTATTCCTTAATGATGCGGCGGCGCATCTTC 540

QY 481 AAGCGATGCGCTTGGGCTTTGTATTCCTTAATGATGCGGCGGCGCATCTTC 540  
 DB 481 AAGCGATGCGCTTGGGCTTTGTATTCCTTAATGATGCGGCGGCGCATCTTC 540

QY 541 ACAGCTTATGAAATTCGCTTAGAAGGCGTGGTCTGTTCTTATTTCTTAATTTGGTTAT 600  
 DB 541 ACAGCTTATGAAATTCGCTTAGAAGGCGTGGTCTGTTCTTATTTCTTAATTTGGTTAT 600

QY 541 ACAGCTTATGAAATTCGCTTAGAAGGCGTGGTCTGTTCTTATTTCTTAATTTGGTTAT 600  
 DB 541 ACAGCTTATGAAATTCGCTTAGAAGGCGTGGTCTGTTCTTATTTCTTAATTTGGTTAT 600

QY 601 TGTGTAACCTCGTCGCTAGGCGAGGTATCCGACCTGTTTAACTGGAATGAGTACAT 660  
 DB 601 TGTGTAACCTCGTCGCTAGGCGAGGTATCCGACCTGTTTAACTGGAATGAGTACAT 660

QY 601 TGTGTAACCTCGTCGCTAGGCGAGGTATCCGACCTGTTTAACTGGAATGAGTACAT 660  
 DB 601 TGTGTAACCTCGTCGCTAGGCGAGGTATCCGACCTGTTTAACTGGAATGAGTACAT 660

QY 661 CGGCTTCTTGGGATATGCTCGGTAGGCGAGATGTCAGTGGGCTGTTTGGTGGCT 720  
 DB 661 CGGCTTCTTGGGATATGCTCGGTAGGCGAGATGTCAGTGGGCTGTTTGGTGGCT 720

QY 661 CGGCTTCTTGGGATATGCTCGGTAGGCGAGATGTCAGTGGGCTGTTTGGTGGCT 720  
 DB 661 CGGCTTCTTGGGATATGCTCGGTAGGCGAGATGTCAGTGGGCTGTTTGGTGGCT 720

QY 721 CATTTCAATGGGCGAAATCTCTCTTACCTATGATGATCATCGATTTTGTATGATGAT 780  
 DB 721 CATTTCAATGGGCGAAATCTCTCTTACCTATGATGATCATCGATTTTGTATGATGAT 780

QY 721 CATTTCAATGGGCGAAATCTCTCTTACCTATGATGATCATCGATTTTGTATGATGAT 780  
 DB 721 CATTTCAATGGGCGAAATCTCTCTTACCTATGATGATCATCGATTTTGTATGATGAT 780

QY 781 TTGGTCTTACAGCGCGGTTTGTATCAGACCGTGTAGCAGCAAAATAGGATGATGAT 840  
 DB 781 TTGGTCTTACAGCGCGGTTTGTATCAGACCGTGTAGCAGCAAAATAGGATGATGAT 840

QY 781 TTGGTCTTACAGCGCGGTTTGTATCAGACCGTGTAGCAGCAAAATAGGATGATGAT 840  
 DB 781 TTGGTCTTACAGCGCGGTTTGTATCAGACCGTGTAGCAGCAAAATAGGATGATGAT 840

QY 841 GAAACAGATTTTGAATCTTGTACAGGCACTGTCATCAAGGATGTTGGGTTGAAATGA 900  
 DB 841 GAAACAGATTTTGAATCTTGTACAGGCACTGTCATCAAGGATGTTGGGTTGAAATGA 900

QY 841 GAAACAGATTTTGAATCTTGTACAGGCACTGTCATCAAGGATGTTGGGTTGAAATGA 900  
 DB 841 GAAACAGATTTTGAATCTTGTACAGGCACTGTCATCAAGGATGTTGGGTTGAAATGA 900

QY 901 ACGAAGCGGCAAGCGTTGTTGATGATGATGATGATGATGATGATGATGATGATGAT 960  
 DB 901 ACGAAGCGGCAAGCGTTGTTGATGATGATGATGATGATGATGATGATGATGATGAT 960

DB 901 ACGAAGCGGCAAGCGTTGTTGATGATGATGATGATGATGATGATGATGATGATGAT 960  
 QY 901 ACGAAGCGGCAAGCGTTGTTGATGATGATGATGATGATGATGATGATGATGATGAT 960

QY 961 CAATCAGTTTCTCTAGAGCTACACGCGAAGTTTGGAAAGCTGCGGAGCGAGTT 1020  
 DB 961 CAATCAGTTTCTCTAGAGCTACACGCGAAGTTTGGAAAGCTGCGGAGCGAGTT 1020

QY 961 CAATCAGTTTCTCTAGAGCTACACGCGAAGTTTGGAAAGCTGCGGAGCGAGTT 1020  
 DB 961 CAATCAGTTTCTCTAGAGCTACACGCGAAGTTTGGAAAGCTGCGGAGCGAGTT 1020

QY 1021 GCTCGGCTAATATTCGAGTTACGATTAATGCGCGGATTTTCCCAATTAAGTACCAAAAC 1080  
 DB 1021 GCTCGGCTAATATTCGAGTTACGATTAATGCGCGGATTTTCCCAATTAAGTACCAAAAC 1080

QY 1021 GCTCGGCTAATATTCGAGTTACGATTAATGCGCGGATTTTCCCAATTAAGTACCAAAAC 1080  
 DB 1021 GCTCGGCTAATATTCGAGTTACGATTAATGCGCGGATTTTCCCAATTAAGTACCAAAAC 1080

QY 1081 CTGGAGTCTAATGCGCAATTTTAAACCAAGCATGCTCAACATCTTACCGTAAAGTGA 1140  
 DB 1081 CTGGAGTCTAATGCGCAATTTTAAACCAAGCATGCTCAACATCTTACCGTAAAGTGA 1140

QY 1081 CTGGAGTCTAATGCGCAATTTTAAACCAAGCATGCTCAACATCTTACCGTAAAGTGA 1140  
 DB 1081 CTGGAGTCTAATGCGCAATTTTAAACCAAGCATGCTCAACATCTTACCGTAAAGTGA 1140

QY 1141 GGAATGACATGGAGCGCGTATGAGTGTCAAGGTATGAGCTTGAAGCTGATGATGATG 1200  
 DB 1141 GGAATGACATGGAGCGCGTATGAGTGTCAAGGTATGAGCTTGAAGCTGATGATGATG 1200

QY 1141 GGAATGACATGGAGCGCGTATGAGTGTCAAGGTATGAGCTTGAAGCTGATGATGATG 1200  
 DB 1141 GGAATGACATGGAGCGCGTATGAGTGTCAAGGTATGAGCTTGAAGCTGATGATGATG 1200

QY 1201 TCATATTTGACCAAGTTGAAAGATTTGATGATTTGAGCCGTGCGTTGATGACCGAGG 1260  
 DB 1201 TCATATTTGACCAAGTTGAAAGATTTGATGATTTGAGCCGTGCGTTGATGACCGAGG 1260

QY 1201 TCATATTTGACCAAGTTGAAAGATTTGATGATTTGAGCCGTGCGTTGATGACCGAGG 1260  
 DB 1201 TCATATTTGACCAAGTTGAAAGATTTGATGATTTGAGCCGTGCGTTGATGACCGAGG 1260

QY 1261 TGAATTTCTTAACCTTACCAATTCGCGGTGAATTTTCAATGCGGCTTGTGCGCTTGCAT 1320  
 DB 1261 TGAATTTCTTAACCTTACCAATTCGCGGTGAATTTTCAATGCGGCTTGTGCGCTTGCAT 1320

QY 1261 TGAATTTCTTAACCTTACCAATTCGCGGTGAATTTTCAATGCGGCTTGTGCGCTTGCAT 1320  
 DB 1261 TGAATTTCTTAACCTTACCAATTCGCGGTGAATTTTCAATGCGGCTTGTGCGCTTGCAT 1320

QY 1321 GTAACGCAATATTTTCAATGCTGGGGGATACCTGTGATCTCAACAGTACTACAGCTTC 1380  
 DB 1321 GTAACGCAATATTTTCAATGCTGGGGGATACCTGTGATCTCAACAGTACTACAGCTTC 1380

QY 1321 GTAACGCAATATTTTCAATGCTGGGGGATACCTGTGATCTCAACAGTACTACAGCTTC 1380  
 DB 1321 GTAACGCAATATTTTCAATGCTGGGGGATACCTGTGATCTCAACAGTACTACAGCTTC 1380

QY 1381 ATGTGATGCTGCTTGGGGTGAATTTTCAACATGATGCGAGTTTATGTTCTTGGCCT 1440  
 DB 1381 ATGTGATGCTGCTTGGGGTGAATTTTCAACATGATGCGAGTTTATGTTCTTGGCCT 1440

QY 1381 ATGTGATGCTGCTTGGGGTGAATTTTCAACATGATGCGAGTTTATGTTCTTGGCCT 1440  
 DB 1381 ATGTGATGCTGCTTGGGGTGAATTTTCAACATGATGCGAGTTTATGTTCTTGGCCT 1440

QY 1441 GATGCGACAGATCACAGGGGAAAAGCCGGCTTGGCGTATCAACAGATCTGATGCGCA 1500  
 DB 1441 GATGCGACAGATCACAGGGGAAAAGCCGGCTTGGCGTATCAACAGATCTGATGCGCA 1500

QY 1441 GATGCGACAGATCACAGGGGAAAAGCCGGCTTGGCGTATCAACAGATCTGATGCGCA 1500  
 DB 1441 GATGCGACAGATCACAGGGGAAAAGCCGGCTTGGCGTATCAACAGATCTGATGCGCA 1500

QY 1501 CATTTACCAAGATCAACTCGAATTTGATGCGGATGCGAGCTTAAACGTTGAGCCATTC 1560  
 DB 1501 CATTTACCAAGATCAACTCGAATTTGATGCGGATGCGAGCTTAAACGTTGAGCCATTC 1560

QY 1501 CATTTACCAAGATCAACTCGAATTTGATGCGGATGCGAGCTTAAACGTTGAGCCATTC 1560  
 DB 1501 CATTTACCAAGATCAACTCGAATTTGATGCGGATGCGAGCTTAAACGTTGAGCCATTC 1560

QY 1561 AGCGCTCAGTTCATATCAATCCAAAGATTAAACATGCGAGATTTTGAACCTTGGGT 1620  
 DB 1561 AGCGCTCAGTTCATATCAATCCAAAGATTAAACATGCGAGATTTTGAACCTTGGGT 1620

QY 1561 AGCGCTCAGTTCATATCAATCCAAAGATTAAACATGCGAGATTTTGAACCTTGGGT 1620  
 DB 1561 AGCGCTCAGTTCATATCAATCCAAAGATTAAACATGCGAGATTTTGAACCTTGGGT 1620

QY 1621 CATTTGATGATTTTGAAGCTGACCGGATATGATTTTCAAGATCTTCAATACCGCTT 1680  
 DB 1621 CATTTGATGATTTTGAAGCTGACCGGATATGATTTTCAAGATCTTCAATACCGCTT 1680

QY 1621 CATTTGATGATTTTGAAGCTGACCGGATATGATTTTCAAGATCTTCAATACCGCTT 1680  
 DB 1621 CATTTGATGATTTTGAAGCTGACCGGATATGATTTTCAAGATCTTCAATACCGCTT 1680

QY 1681 TTCAGTCTAATCCGATATTCAGGCGGATGCGGATGATGATGATTTTATTAATAAAGCTCC 1740  
 DB 1681 TTCAGTCTAATCCGATATTCAGGCGGATGCGGATGATGATGATTTTATTAATAAAGCTCC 1740

QY 1681 TTCAGTCTAATCCGATATTCAGGCGGATGCGGATGATGATGATTTTATTAATAAAGCTCC 1740  
 DB 1681 TTCAGTCTAATCCGATATTCAGGCGGATGCGGATGATGATGATTTTATTAATAAAGCTCC 1740

QY 1741 CGAAGGTGCGGAGCTTTTATTAACAGATGATGCTTAAACGCTTAAAGCGTTAGGCGAAG 1800  
 DB 1741 CGAAGGTGCGGAGCTTTTATTAACAGATGATGCTTAAACGCTTAAAGCGTTAGGCGAAG 1800

QY 1741 CGAAGGTGCGGAGCTTTTATTAACAGATGATGCTTAAACGCTTAAAGCGTTAGGCGAAG 1800  
 DB 1741 CGAAGGTGCGGAGCTTTTATTAACAGATGATGCTTAAACGCTTAAAGCGTTAGGCGAAG 1800

QY 1801 AATGCTGCGGGGATGACGACAAACACCCCAATTAAGTAACTCAACACCACTTTTCT 1860  
 DB 1801 AATGCTGCGGGGATGACGACAAACACCCCAATTAAGTAACTCAACACCACTTTTCT 1860

QY 1801 AATGCTGCGGGGATGACGACAAACACCCCAATTAAGTAACTCAACACCACTTTTCT 1860  
 DB 1801 AATGCTGCGGGGATGACGACAAACACCCCAATTAAGTAACTCAACACCACTTTTCT 1860

QY 1861 CTTAACAACCAAGTTGATGATGAGTGTGACCTTATTAAGCGGTTGCGGTTAGAAAGG 1920  
 DB 1861 CTTAACAACCAAGTTGATGATGAGTGTGACCTTATTAAGCGGTTGCGGTTAGAAAGG 1920

QY 1861 CTTAACAACCAAGTTGATGATGAGTGTGACCTTATTAAGCGGTTGCGGTTAGAAAGG 1920  
 DB 1861 CTTAACAACCAAGTTGATGATGAGTGTGACCTTATTAAGCGGTTGCGGTTAGAAAGG 1920

QY 1921 AATACCGTAAATCAAGACCGTAGCCATCAAGTTTAAAGCTTAAAGTCAACAGGCAATTG 1980  
 DB 1921 AATACCGTAAATCAAGACCGTAGCCATCAAGTTTAAAGCTTAAAGTCAACAGGCAATTG 1980

QY 1921 AATACCGTAAATCAAGACCGTAGCCATCAAGTTTAAAGCTTAAAGTCAACAGGCAATTG 1980  
 DB 1921 AATACCGTAAATCAAGACCGTAGCCATCAAGTTTAAAGCTTAAAGTCAACAGGCAATTG 1980

QY 1981 CAGACCAACACGCGAACTCAACAGACAGCGGTTCCGCGACAGACGATTAATCA 2040  
 DB 1981 CAGACCAACACGCGAACTCAACAGACAGCGGTTCCGCGACAGACGATTAATCA 2040



[illegible][illegible]

```

QY 1903 CAGTTGCGTAAGAAAGAAATACCGTAATCAAGACCGTAAGCCATCAAGTTAAAGCTTAA 1962
DB 1029 CAGTTGCGTAAGAAAGAAATACCGTAATCAAGACCGTAAGCCATCAAGTTAAAGCTTAA 970
QY 1963 GTGACACGAGCAATTTGAGAGCAAAACCGCAAACTCAACGAGACAGCGGTTGCGGAC 2022
DB 969 GTGACACGAGCAATTTGAGAGCAAAACCGCAAACTCAACGAGACAGCGGTTGCGGAC 910
QY 2023 GAGCAGAGCAATTAATCAAGTGCATATGTTCCGCACTTAAGATGATGATGATTTCAAG 2082
DB 909 GAGCAGAGCAATTAATCAAGTGCATATGTTCCGCACTTAAGATGATGATGATTTCAAG 850
QY 2083 CACTTCAGCAGCGCAGAGCCCAAGAGAGAACCAATTAAGCTGTTGTGTGTGATGAGA 2142
DB 849 CACTTCAGCAGCGCAGAGCCCAAGAGAGAACCAATTAAGCTGTTGTGTGTGATGAGA 790
QY 2143 TTGAATTAATACCGTAACCACTGTACTGAGCAATACCGTGTATGATGATGATTTGCGC 2202
DB 789 TTGAATTAATACCGTAACCACTGTACTGAGCAATACCGTGTATGATGATGATTTGCGC 730
QY 2203 CGCATTTTGTAGATTTTCAAGTGGGCGCAACCAATCACTTATCATGATTTGCGCAT 2262
DB 729 CGCATTTTGTAGATTTTCAAGTGGGCGCAACCAATCACTTATCATGATTTGCGCAT 670
QY 2263 CACCGTAATGCGAGCAGAAATGCTGCAATACCAATACGATTAAGTGCACACCAACGAA 2322
DB 669 CACCGTAATGCGAGCAGAAATGCTGCAATACCAATACGATTAAGTGCACACCAACGAA 610
QY 2323 AGTATTAATCCATTAACGAAAGCTGGGTTTCAAGCCCTGTGATGACAGTTTGTATTCG 2382
DB 609 AGTATTAATCCATTAACGAAAGCTGGGTTTCAAGCCCTGTGATGACAGTTTGTATTCG 550
QY 2383 TTGTATTAATCAAAACCTTCAATGCTCATATGCGCAGTGCAGAGAGAGGCGAAACGAG 2442
DB 549 TTGTATTAATCAAAACCTTCAATGCTCATATGCGCAGTGCAGAGAGAGGCGAAACGAG 490
QY 2443 CCAGTGTGAGACTTCTCTTAATATGCAAAACATATTTCTAGAGTGAAGATGACGAC 2502
DB 489 CCAGTGTGAGACTTCTCTTAATATGCAAAACATATTTCTAGAGTGAAGATGACGAC 430
QY 2503 CGCGAGAGATGAAAAAATCGTGTATGCTGCGCACTGCGGAAAGCAACGCGGAACTTTC 2562
DB 429 CGCGAGAGATGAAAAAATCGTGTATGCTGCGCACTGCGGAAAGCAACGCGGAACTTTC 370
QY 2563 TTTAACGCGCATATGCGCAAGGCTGACGAGATGATGTCACATGTAACCAATATTCG 2622
DB 369 TTTAACGCGCATATGCGCAAGGCTGACGAGATGATGTCACATGTAACCAATATTCG 310
QY 2623 ACCCATCAACATAGAAATGCGGTTTCAACCGGTAAACCAACGCAACGCAACAT 2682
DB 309 ACCCATCAACATAGAAATGCGGTTTCAACCGGTAAACCAACGCAACGCAACAT 250
QY 2683 AATAGAGTCAACCGTCTTGAAGATTTGAATCAAGTGCCTTGCATTAACCAATCA 2742
DB 249 AATAGAGTCAACCGTCTTGAAGATTTGAATCAAGTGCCTTGCATTAACCAATCA 190
QY 2743 TCTGTCAATTTGGTGGAGCAAAATTCAAATAGAACTTTGGCTGATTCGCGGTTGCCA 2802
DB 189 TCTGTCAATTTGGTGGAGCAAAATTCAAATAGAACTTTGGCTGATTCGCGGTTGCCA 130
QY 2803 TTTAAACCGCTGCGCAACCATGCGCACTGCAAGAAATAGTAATAACAGATGAAGCA 2862
DB 129 TTTAAACCGCTGCGCAACCATGCGCACTGCAAGAAATAGTAATAACAGATGAAGCA 70
QY 2863 GTTTGCGCAACGATGAGCTTTGTGTGTCAAGCAATCGCGCTGCAG 2909
DB 69 GTTTGCGCAACGATGAGCTTTGTGTGTGTCAAGCAATCGCGCTGCAG 23

```

RESULT 4  
AA224801  
ID AA224801 standard; DNA; 852 BP.  
XX

```

AC AA224801;
DB 07-DEC-1999 (first entry)
XX
DE Vibrio cholerae thya coding region sequence.
XX
KW Wild type; thya; enzyme; thymidilate synthase; non-toxicogenic; mutation;
KM attenuation; immunization; cholera; hemagglutinin protease; biosafety;
KM auxotrophism; replication; vaccine; ss.
XX
OS Vibrio cholerae.
XX
PN W0935271-A2.
XX
PD 15-JUL-1999.
XX
PF 30-DEC-1998; 98WO-CU000008.
XX
PR 30-DEC-1997; 97CU-00000142.
XX
PA (NAIIN-) CENT NACIONAL INVESTIGACIONES CIENTIFICAS.
XX
PI Campos Gomez J, Fando Calzada RA, Rodriguez Gonzalez BL,
PI Ledon Perez TY, Valle Diaz E, Silva Cabrera AD, Benitez Robles JA;
DR MPI; 1999-430398/36.
XX
DR P-PSDB; AA226895.
XX
PT Producing strains of Vibrio cholerae with inactivated gene for
PT hemagglutinin protease, useful in vaccines against cholera.
XX
PS Claim 18; Page 30; 30pp; Spanish.
XX
CC This sequence represents the wild type coding region of the thya gene
CC from Vibrio cholerae, which encodes the enzyme thymidilate synthase. The
CC invention relates to the production, from a non-toxicogenic strain of V.
CC cholerae, of attenuated strains of V. cholerae suitable for immunization
CC against cholera. The attenuation comprises inactivating the gene for
CC hemagglutinin protease (HPI), either by deletion, insertion or some other
CC defined and irreversible genetic manipulation. Additional biosafety of
CC the attenuated strains is generated by mutating the thya gene, resulting
CC in auxotrophic mutants unable to replicate in the environment. The new
CC strains are used to produce anticholera vaccines
XX
SQ Sequence 852 BP; 217 A; 177 C; 216 G; 242 T; 0 U; 0 Other;
Query Match 29.0%; Score 842.4; DB 2; Length 852;
Best Local Similarity 99.3%; Pred. No. 1.7e-257;
Matches 846; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 839 GTGAACAGATATTTGATCTTTGTGACGCGCATCGCATCAAGATGTTGGTTGAAT 898
DB 1 GTGAACAGATATTTGATCTTTGTGACGCGCATCGCATCAAGATGTTGGTTGAAT 60
QY 899 GAACGAAACGCGCAACGTTGTTGAATGATTAATGCGATTTGACCTAGCATGTGGC 958
DB 61 GAACGAAACGCGCAACGTTGTTGAATGATTAATGCGATTTGACCTAGCATGTGGC 120
QY 959 AACATCAAGTTTCTCTAGTGAATACAGCAAGATTTTGGAAAGCTGCGTACCGAG 1018
DB 121 AACATCAAGTTTCTCTAGTGAATACAGCAAGATTTTGGAAAGCTGCGTACCGAG 180
QY 1019 TTGCTCGGTATTTCTGTGATTAATGAGTGGGAGATTTTGGCAATTAAGTACCAA 1078
DB 181 TTGCTCGGTATTTCTGTGATTAATGAGTGGGAGATTTTGGCAATTAAGTACCAA 240
QY 1079 ACCTGGATGCTAATGCAATTTAAACCAAGATGGCTCAACATCTTTACCGTAAAGT 1138
DB 241 ACCTGGATGCTAATGCAATTTAAACCAAGATGGCTCAACATCTTTACCGTAAAGT 300
QY 1139 GAGATGACATGGAACGCGTGTATGATGTTTCAAGGATGAGCTTGGCTAAGCTGATGT 1198
DB 301 GAGATGACATGGAACGCGTGTATGAGTTCAGGATGAGCTTGGCTAAGCTGATGT 360

```

QY 1199 GGTGATATTGACCAAGTTGAAAAAGATTGTTGATGATTTGAGCCGTGGCGTTGATGACCGA 1258  
 DB 361 GGTGATATTGACCAAGTTGAAAAAGATTGTTGATGATTTGAGCCGTGGCGTTGATGACCGA 420  
 QY 1259 GGTGAAATTTCTTAACCTTCACAAATCCGGGTGAATTTGACATGGGGGTGTTGGCCCTTGC 1318  
 DB 421 GGTGAAATTTCTTAACCTTCACAAATCCGGGTGAATTTGACATGGGGGTGTTGGCCCTTGC 480  
 QY 1319 ATGTACAGCCATCATTTTTCATTTGCTGGGGGATACCTTGTATCTGAACAGTACTGAGCGT 1378  
 DB 461 ATGTACAGCCATCATTTTTCATTTGCTGGGGGATACCTTGTATCTGAACAGTACTGAGCGT 540  
 QY 1379 TCATGTGATGTGCTTGGGGGTGTAATTTCAACATGTGACAGTTTATGTTCCTTGG 1438  
 DB 541 TCATGTGATGTGCTTGGGGGTGTAATTTCAACATGTGACAGTTTATGTTCCTTGG 600  
 QY 1439 CTGATGGCAGATCAACAGGAAAAAGCCGGCTTGGCGTATCAACAAGTGTCAATGCG 1498  
 DB 601 CTGATGGCAGATCAACAGGAAAAAGCCGGCTTGGCGTATCAACAAGTGTCAATGCG 660  
 QY 1499 CACATTTTACCAAGATCACTCGAATTTGATGCGCGATGTCAGACTTAAACGTGAGCCATTG 1558  
 DB 661 CACATTTTACCAAGATCACTCGAATTTGATGCGCGATGTCAGACTTAAACGTGAGCCATTG 720  
 QY 1559 CCAGGCGCTCAGTTCATATCAATCAATCAAAAGATTAAACATGTCAGAGATTGGAACTTGG 1618  
 DB 721 CCAGGCGCTCAGTTCATATCAATCAATCAAAAGATTAAACATGTCAGAGATTGGAACTTGG 780  
 QY 1619 GTCACTTTGATGATTTTGAAGTCAACCGGATATCATGTTCCAGATCCATTAAATACCG 1678  
 DB 781 GTCACTTTGATGATTTTGAAGTCAACCGGATATCATGTTCCAGATCCATTAAATACCG 840  
 QY 1679 TTTTCAGTCTAA 1690  
 DB 841 TTTTCAGTCTAA 852

RESULT 5  
 ID ACAA53066  
 AC53066 standard; DNA; 852 BP.  
 AC ACAA53066;  
 DT 19-JUN-2003 (first entry)  
 DE Prokaryotic essential gene #34723.  
 KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
 KM drug design; gene.  
 OS Vibrio cholerae.  
 FN WO200277183-A2.  
 PD 03-OCT-2002.  
 PE 21-MAR-2002; 2002MO-US009107.  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 PA (ELIT-) ELITRA PHARM INC.  
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 DR Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 P-WPI; 2003-029926/02.  
 P-PSDB; AB049196.

PT New antisense nucleic acid, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 PS Claim 14; SEQ ID NO 40936; 1766bp; English.  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-regulated gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pce\_sequences  
 CC XX

Query Match 29.0%; Score 842.4; DB 7; Length 852;  
 Best Local Similarity 99.3%; Pred. No. 1.7e-257;  
 Matches 846; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 QY 839 GTGAAACAGTATTGATCTTTGTCAGCGCATCGCATCAAGGTGTTGGTGAAT 898  
 DB 1 GTGAAACAGTATTGATCTTTGTCAGCGCATCGCATCAAGGTGTTGGTGAAT 60  
 QY 899 GAACGACGGCAAGCGTTGTTGACTGTGATTAATGCCGATTGACCTACGATGTGGC 958  
 DB 61 GAACGACGGCAAGCGTTGTTGACTGTGATTAATGCCGATTGACCTACGATGTGGC 120  
 QY 959 AACCAATCACTTCTCTATGATGATCAACGCAAGTTTGGAAAGCGCGTGACCGAG 1018  
 DB 121 AACCAATCACTTCTCTATGATGATCAACGCAAGTTTGGAAAGCGCGTGACCGAG 180  
 QY 1019 TTGCTCGGCTAATTCGTGTTACGATATGCGCGGATTTTCCGCAATTAGTAAACAA 1078  
 DB 181 TTGCTCGGCTAATTCGTGTTACGATATGCGCGGATTTTCCGCAATTAGTAAACAA 240  
 QY 1079 AACTGGAGTGTAAATGCCAATTTAAACCAAGCATGCTCAACAATCTTACCGTTAAGT 1138  
 DB 241 AACTGGAGTGTAAATGCCAATTTAAACCAAGCATGCTCAACAATCTTACCGTTAAGT 300  
 QY 1139 GAGGATGACATGGGACGCGGTGATAGTTCAGGGTAAAGCTTGGCGTAAACCGATGCT 1198  
 DB 301 GAGGATGACATGGGACGCGGTGATAGTTCAGGGTAAAGCTTGGCGTAAACCGATGCT 360  
 QY 1199 GGTGATATTGACCAAGTTGAAAAAGATTGTTGATGATTTGAGCCGTGGCGTTGATGACCGA 1258  
 DB 361 GGTGATATTGACCAAGTTGAAAAAGATTGTTGATGATTTGAGCCGTGGCGTTGATGACCGA 420  
 QY 1259 GGTGAAATTTCTTAACCTTCACAAATCCGGGTGAATTTGACATGGGGGTGTTGGCCCTTGC 1318





RESULT 8  
 ID ACA42822 standard; DNA; 852 BP.  
 AC ACA42822;  
 DT 19-JUN-2003 (first entry)  
 DE Prokaryotic essential gene #24479.  
 KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
 KW drug design; gene.  
 OS Pasteurella multocida.  
 PN WO20027183-A2.  
 PD 03-OCT-2002.  
 PF 21-MAR-2002; 2002WO-US009107.  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362659P.  
 PA (ELIT-) ELITRA PHARM INC.  
 PI Wang L, Zamudio C, Malone C, Haseelbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX P-PSDB; AB038952.  
 DR MPI; 2003-029926/02.  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 PS Claim 14; SEQ ID NO 30692; 1766bp; English.  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 852 BP; 247 A; 201 C; 181 G; 223 T; 0 U; 0 Other;  
 Query Match 17.4%; Score 505.4; DB 7; Length 852;  
 Best Local Similarity 74.6%; Pred. No. 6.8e-150;  
 Matches 635; Conservative 0; Mismatches 216; Indels 0; Gaps 0;  
 QY 840 TGAACAGATTTAGATCTTTGTCAGCGCATGCGATCAAGGTGTTGGTGAAGAATG 899  
 DB 2 TGAACACATTTTAGACTTTGACCTTGCACAGATGATGATGAAGTGTGGTGAAGAATG 61  
 QY 900 AACGAACGGGCAAGCGTTGTTGACTGTGATTAATCCGATTTGACCTACGATGTGGCA 959  
 DB 62 CACGACCTGGCAAAACGTTGCTGACCGGTATGATGCGGATCTCACTACGATGTGGCA 121  
 QY 960 ACATACGTTTCTCTAGAGACTTACACGCAAGATTTTGAAGTGCCTGACCGGACT 1019  
 DB 122 ATTAACCAATTTCCGTTGATCACACACGCAAAAGCTATTGGAAGCGCAATTGCAAA 181  
 QY 1020 TGCTGGGCTATATATCGTGGTTACGATTAATGCGGCGGATTTTGGCAATTAGGTACCAAA 1079  
 DB 182 TTTTAGTTTATATCCGTGGCTATGACAAATGCGGCAATTTCCGCAACTCGGACCAAAA 241  
 QY 1080 CCTGGGATGCTAATGCAATTTAAACCAAGCATGCTCAACATCTTAACGTTAAAGGTG 1139  
 DB 242 CCGGGATGCAATGCTAATGAAAACAGTGATGCTCAATATCCGACCGCAAGGCA 301  
 QY 1140 AGATGACATGGGAGCGGCTGATGCTGTTCAAGGTAGAGCTTGGGCTAGCGTGTG 1199  
 DB 302 CCGATGATATGGGACGCTGTTATGGCGTAAAGGGGCTGATGCGCAACCGAAGCGTG 361  
 QY 1200 GTCATATTACCAAGTTGAAAAGATTTGATGATTTGAGCCGTGGCTGTATGACCGAG 1259  
 DB 362 AAACGCTCATTAATTTGCGCAAAATGCTCAATATATGCGCAACGGCATTTGACCGTG 421  
 QY 1260 GTGAATTTCTTAATCTTCAAAATCCGGGTGAATTTTCAATGGGAGTGTTCGCCCTTGA 1319  
 DB 422 GTGAATATATGACCTTTCTTCAACCCGGGTAATTTGAATGGGTTCCATACGCTTGA 481  
 QY 1320 TGTACAGCATATTTTCTATGCTGCGGAGTAACCTTGTATCTCAACAGTACTCAGCGTT 1379  
 DB 482 TGCACACCCACACCTTCTCATTAATCTCGGTGATTAACCTCTATCTCAACAGTATCAGCGTT 541  
 QY 1380 CATTGATGTCGCTTGGGTTGAAATTTCAACATGCTGAGGTTATGTTCTTGGCG 1439  
 DB 542 CTGTGATGTCGCTTGGCTTGAATTTCAACAAATTTCAAGTTCACTTTCTTGGCT 601  
 QY 1440 TGATGCAACAGATCAACAGGAAAAAGCCGCGCTTGGCTGATCAACAGATGCTCAATGCC 1499  
 DB 602 TGATGCAACAAATTAACGGGCAAAAAAGCGGTAAAGCGTAAACAAATCATCAATGAC 661  
 QY 1500 AATTTTACCAAGATCACTGAAATGATGCGGATGTCAGTCAAAACGTGACCATTC 1559  
 DB 662 AATTTTATTAAGATCAACTGATCTCATGAAAAAATGTGCAATTTAAACGTGACCCCTTGC 721  
 QY 1560 CAGCGCTCAGTTCCATATCATTCCAAGATTTAAACCTGAGATTTGGAACCTTGG 1619  
 DB 722 CATTGCAACAAATTAACATTAACCCGGAATTTAAACCTGCAAGATTTGAACCTTGG 781  
 QY 1620 TCATTTGGATGATTTTGAAGTCAACCGGATATCAAGTTCAAGTCTTATTCATCCGT 1679  
 DB 782 TGACATATGATATTTTAAAGTCAACGGGCTATCAATGCAACATGCAATTAATATCTT 841  
 QY 1680 TTTCAAGTCTAA 1690  
 DB 842 TCTCTGTGTA 852  
 RESULT 9  
 ID ACA34292 standard; DNA; 852 BP.  
 AC ACA34292;  
 XX ftp.wipo.int/pub/published\_pct\_sequences

XX 19-JUN-2003 (first entry)  
 XX Prokaryotic essential gene #15949.  
 DE  
 XX Antisense; de; prokaryotic essential gene; cell proliferation;  
 KM drug design; gene.  
 XX  
 OS Haemophilus influenzae.  
 XX  
 PN WO20027183-A2.  
 PD 03-OCT-2002.  
 XX  
 PF 21-MAR-2002; 2002WO-US009107.  
 XX  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW,  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH,  
 XX MPI; 2003-029926/02.  
 DR P-P8DB; ABU30422.  
 XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 BS Claim 14; SEQ ID NO 22162; 1766bp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 852 BP; 246 A; 152 C; 193 G; 261 T; 0 U; 0 Other;

Query Match 15.9%; Score 462.2; DB 7; Length 852;  
 Best Local Similarity 71.4%; Pred. No. 4.2e-136;  
 Matches 608; Conservative 0; Mismatches 243; Indels 0; Gaps 0;

QY 840 TGAACAGATATTAGATCTTTGTGACGCGCATCGTCATCAAGTGTGTTGGTTGAATAATG 899  
 |||||  
 DB 2 TGAAGCAATATCTTGAGCTTTGTGCGCGCATTTGTTAGTGAAGGGAATGGGTCTTAATG 61  
 QY 900 AACGACGGGCAACCGTTGTTGACTGTGATTAATGCCGATTTGACCTACGATGCGGCA 959  
 |||||  
 DB 62 AACGTACAGGTACCATTTGGCTCAGCGGTATTAATGCAAGTTTGAATGATGATGTGGGA 121  
 QY 960 ACAATCAGTTCTCTGATGACTACACGACAGATTTTGGAAAGTGCCTAGCCGAGT 1019  
 |||||  
 DB 122 ATTAATCAATTTCCGCTGATTAATACCGGTAAAGTTATGGAAGGCGCATTTGCTGAT 181  
 QY 1020 TGTCTGGCTAATATTCGTGTTAGCATTAATGCGGCGGATTTTGGCAATTTAGTACCAAAA 1079  
 |||||  
 DB 182 TTTTAGGTTATATTCGTGATGATGATGCAATGCGCTGATTTCCGCGCACTTGGCAGAAA 241  
 QY 1080 CCTGGATGCTTAATGCCAATTTAAACAGATGCGTCAACAAATCTTACGTTAAAGGTG 1159  
 |||||  
 DB 242 CTTGGGATGCTTAATGCCAATGAAATGCAAGCTTGGCTTGCATTAATCCGATGTAAGGCG 301  
 QY 1140 AGATGACATGAGGACGCGTGTATGTTGTTGAGGTTAGAGCTTGGGCTTAAGCGTATGAGTG 1159  
 |||||  
 DB 302 TTGATGATATGGGCGCGTATATGTGTGCAAGGACAGCATGCGTAAAGCTTAATGAG 361  
 QY 1200 GTCATATTGACCACTTGAAAAAGATTGTGATGATTTGAGCGCGTGTGATGACCGAG 1259  
 |||||  
 DB 362 AAATCATGATCAGCTACGTAAATTTGTTAATTAATTAATTAATTAATTAATTAATTAAT 421  
 QY 1260 GTGAAATTTCTTAATCTTCAATATCCGGGTAAATTTCAATGGGAGTTTGGCCCTTGCA 1319  
 |||||  
 DB 422 GAGGATTTTAACTTTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 481  
 QY 1320 TGTACAGCATCATTTTTCATTTGCTGGGGGATACCTTGATCTCAACAGTACTCAGCGTT 1379  
 |||||  
 DB 482 TGCATAGCATATCTTTTCTTGTGGCGATCTTACATCTTACATCTTACATCTTACATCTT 541  
 QY 1380 CATGTGATGTCCTTGGGGTGAATTTCAACATGTGTCAGGTTTATGTCTTGGCG 1439  
 |||||  
 DB 542 CCGTGAATGTTCCGCTGGATGATTAATTTCAATCAATTTCAAGTGTTAATCTTAAAGC 601  
 QY 1440 TGATGGCAGCATACACAGGAAAAAGCCGGCTTGGGTATCAACAAGTGTCAATGGCG 1459  
 |||||  
 DB 602 TTATGGCTCAGATACCGGCAAAAAAGCCGGCAAGGCAATATCAATTAATTTGTGATGGC 661  
 QY 1500 ACATTACCAAGATCACTGATGATGAGCGGATGTCAGCTAAACGTAGGCATTTCC 1559  
 |||||  
 DB 662 ATATTATGAAGATCAGCTTGAATGATGCTGACGTACCACTTAACGCAAGCTTTCC 721  
 QY 1560 CAGGCGCTCAGTTCATATCAATCCAAAGATTAAACACTGACAGATTTGGAAACTTGGG 1619  
 |||||  
 DB 722 CATTAACCAAAATTTGAATTAATCAAGATTAATAAAAGCTTGAAGATTTGGAAACTTGGG 781  
 QY 1620 TCACCTTGGATGATTTTGACGTACCGGATATCAGTTTCACAGATCTTAAATACCGGT 1679  
 |||||  
 DB 782 TCACGATGAGATTTTAAAGTGTGCTATCAATCCAGAAACCAATTAATATATCTT 841  
 QY 1680 TTTCAGTCTAA 1690  
 |||||  
 DB 842 TTTCGTTTAA 852  
 |||||  
 RESULT 10  
 ID ACA44455 standard; DNA; 849 BP.  
 AC ACA44455;  
 XX  
 XX 19-JUN-2003 (first entry)  
 XX  
 XX Prokaryotic essential gene #26112.  
 DE Antisense; de; prokaryotic essential gene; cell proliferation;  
 XX

KW drug design; gene.  
 XX  
 OS Proteus sp.  
 XX  
 PN WO200277183-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 21-MAR-2002; 2002MO-US009107.  
 XX  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362659P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW,  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Foreyth RA, Xu HH;  
 XX  
 DR WPI; 2003-029926/02.  
 DR P-PSDB; AB040585.  
 XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 PS Claim 14; SEQ ID NO 32325; 1766bp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 849 BP; 263 A; 155 C; 189 G; 242 T; 0 U; 0 Other;

Query Match 13.3%; Score 403.8; DB 7; Length 849;  
 Best Local Similarity 67.3%; Pred. No. 1.9e-117;  
 Matches 570; Conservative 0; Mismatches 277; Indels 0; Gaps 0;

QY 840 TGAACAGATTTTAAATCTTTGTCAGCGCATCGTCGATCAAGGTGTTGGCTGAAATG 899  
 DB 2 TGAAGCAGATTCGCGATTCGTCACGCAATTCATGATGAGGTGATGATGATTAATA 61  
 QY 900 AACGAAGGCGCAAGCGTTGTTGACTGTGATTAATGCGCATTTGACCTACGATGTGGCA 959

DB 62 AACGACAGAAACCCGTTGTTTAAACGTCATAAATGCCATCGAATATGATGTTGCCA 121  
 QY 960 ACAATGAGTTTCTCTAGTACATACACGCAAGTTTTGGAAGTCGCGTACCGACT 1019  
 DB 122 ATBACCAATTTCCACTCATACACGCGTAAAGTTTAAACAAGCGGATTCGGGAAC 181  
 QY 1020 TCGTCGGCTATATTCGTGGTTCAGTAAATGCGCGGATTTTGGCCATTAAGTACCAAA 1079  
 DB 182 TATTAAGGTAATACGTGTTATGATTAATGACGACCAATTTGTGCTATATGGCTGAATA 241  
 QY 1080 CCTGGAGTCTAATGCAATTTAAACCAAGCATGCTCAACAATCTTAACGTAAGGTG 1139  
 DB 242 CCGTGAACCGCAATGCTAATGAAATAGCGCATGTTAATATCCGATTCGTAAGGCTG 301  
 QY 1140 AGATGACATGAGGACCGGTATGCTTCAAGGTATAGACTTGGCTAAGCCGTAGTGTG 1199  
 DB 302 AAGATGATATGAGCGCGAGTATGCTTCAAGGACGCCAATGCAACGCCCTGATGCT 361  
 QY 1200 GTCATATTAACCAAGTTGAAAGATTTGATGATTTGACCGGTGATGACCGAG 1259  
 DB 362 CGCATCTTATCAATTAAGTAAAGTGTATGATTAATCAATCAATGATTAAGTACGATCG 421  
 QY 1260 GTGAATTTCTTAATCTTCAAAATCCGGGTGAATTTCAATGCGGTGTTTGGCCCTTGA 1319  
 DB 422 GTGAGTATGACTTTTATTAATCCGAGAAACAGCGTTAGGCTTATACGTCATGCA 481  
 QY 1320 TGTACAGCCATATTTTATGCTGCGGAGTAACTTGTATCTCAACAGTACTGACGCT 1379  
 DB 482 TGCATACATATCTTTTATGCTGCGGAGTAACTTATTAATCTCATATCAACGTA 541  
 QY 1380 CATGATATGCGCTTGGGCTTGAATTTCAACATGCTGACGTTATATGTTCTTGGCC 1439  
 DB 542 GCTGTATGATCCATTAAGGTTGAATTTTAACTTAATTAATGCTTATTAATGAGCC 601  
 QY 1440 TGAATGACAGATCAACAGGAAACCCGCGCTTGGCGTATCAACAAGATCGTCAATGCG 1499  
 DB 602 TGGTGGCTCAATCAACAGGCAATTAACAGTAAAGCATTAAGATTAATGATGCTC 661  
 QY 1500 AACTTTACCAATTAATCAATCGAATGATGCGGATGTCAGCTTAAGTACGATTCG 1559  
 DB 662 AACTTTATGAAACCAATTAATCGTATGCTGATGCTGATGCTGATTAATTAAGGCAATTC 721  
 QY 1560 CAGCGCTCAGTTCATATCAATCAAGATTAATAAAGTGAAGTTGGAACCTTGGG 1619  
 DB 722 CATTAAGTACATTAATTAATCAACATTAAGCATTAAGTATTAAGAACTTGGG 781  
 QY 1620 TCACTTTGATGATTTTGAAGTCAACCGATATCAAGTTCAGATTCATTAATCCCGT 1679  
 DB 782 TCAACAGATATGATTAATGATTAAGTATCAAGTATCAAGTATCAAGGATTAATCCCGT 841  
 QY 1680 TTTCACT 1686  
 DB 842 TCAAGCT 848

RESULT 11  
 AA237118  
 ID AA237118 standard; DNA; 807 BP.  
 XX  
 AC AA237118;  
 XX  
 DT 28-JUN-2000 (first entry)  
 XX  
 DE Nucleotide sequence of H.influenzae HI0904 gene.  
 XX  
 KW Genome; mutagenesis; transposon; isolate; locate; essential gene; ss;  
 KW detect; growth; anti-microbial therapy; genomic footprinting.  
 OS Haemophilus influenzae.  
 XX  
 PN WO950402-A1.

PD 07-OCT-1999.  
 XX 26-MAR-1999; 99WO-US006139.  
 XX 27-MAR-1998; 98US-0079770P.  
 PR (HARD ) HARVARD COLLEGE.  
 XX Mekalanos JJ, Akerley B, Rubin E, Camilli A;  
 XX WPI; 1999-620065/53.  
 DR Detection of genes essential for the growth and viability of organism,  
 PT useful as e.g. targets for anti-microbial therapy.  
 XX  
 PS Disclosure; Fig 5F; 62pp; English.  
 CC Nucleotide sequences AA27113-237126 are essential Haemophilus influenzae  
 CC genes that have been identified by the methods of the invention. The  
 CC invention relates to a method for locating essential regions of a portion  
 CC of an organism's genome by: (i) in vitro mutagenising DNA having the  
 CC sequence of the region with a transposon; (ii) identifying cells of the  
 CC organism transformed with the mutagenised DNA of (i); and (iii) locating  
 CC the region by detecting the absence of transposons in the region of the  
 CC mutagenised cells containing the mutagenised DNA. The invention also  
 CC relates to a method for isolating a compound that modulates the  
 CC expression of a nucleic acid sequence operably linked to a gene promoter,  
 CC and a method of identifying a nucleic acid sequence that is essential for  
 CC cell growth or viability. The methods are used to detect genes that are  
 CC essential for the growth and viability of organism. Such genes can be  
 CC used e.g. as targets for anti-microbial therapy  
 CC  
 SQ Sequence 807 BP; 178 A; 139 C; 185 G; 305 T; 0 U; 0 Other;  
 Query Match 10.0%; Score 292.2; DB 2; Length 807;  
 Best Local Similarity 65.0%; Pred. No. 7.8e-82;  
 Matches 467; Conservative 0; Mismatches 243; Indels 9; Gaps 2;

QY 618 TAGGAGCGTATCCGAGCTGTTTATAGCTGATACGATATCCGCTTCTTGAT 677  
 DB 614 TGGCTTCTGTGACAGGTTATTTCTTAATGGTATAGCGCTTCGCTTTATGTGAAAT 673  
 QY 678 ACGTCCGAGCCAGATCCTCAGTTGGCTCTGTTGGGCTTCAATTCAATGGGCAAA 737  
 DB 674 ATGTGCGGAACCTGAAAG-----TTGAAATTTCTTTGGATATTAACAGAGGCAAG 727  
 QY 738 TCCCTCCTTACTATAGTATGATATCGGATATTTGATGATGTTGTCTTACAAGCCG 796  
 DB 728 CCTTTGCTTCCAGATGATATTAATGTGTGCTTTCATTATAGCTTGGGCTTATTCAGCC 786

RESULT 12  
 Continuation (28 of 57) of ACF67367 from base 2700001 (Photorehabus luminescens nucleoti.  
 WP Sequence split into 57 fragments LOCUS ACF67367 Accession ACF67367  
 WP Fragment Name Begin End  
 WP ACF67367\_00 1 110000  
 WP ACF67367\_01 100001 210000  
 WP ACF67367\_02 200001 310000  
 WP ACF67367\_03 300001 410000  
 WP ACF67367\_04 400001 510000  
 WP ACF67367\_05 500001 610000  
 WP ACF67367\_06 600001 710000  
 WP ACF67367\_07 700001 810000  
 WP ACF67367\_08 800001 910000  
 WP ACF67367\_09 900001 1010000  
 WP ACF67367\_10 1000001 1110000  
 WP ACF67367\_11 1100001 1210000  
 WP ACF67367\_12 1200001 1310000  
 WP ACF67367\_13 1300001 1410000  
 WP ACF67367\_14 1400001 1510000  
 WP ACF67367\_15 1500001 1610000  
 WP ACF67367\_16 1600001 1710000  
 WP ACF67367\_17 1700001 1810000  
 WP ACF67367\_18 1800001 1910000  
 WP ACF67367\_19 1900001 2010000  
 WP ACF67367\_20 2000001 2110000  
 WP ACF67367\_21 2100001 2210000  
 WP ACF67367\_22 2200001 2310000  
 WP ACF67367\_23 2300001 2410000  
 WP ACF67367\_24 2400001 2510000  
 WP ACF67367\_25 2500001 2610000  
 WP ACF67367\_26 2600001 2710000  
 WP ACF67367\_27 2700001 2810000  
 WP ACF67367\_28 2800001 2910000  
 WP ACF67367\_29 2900001 3010000  
 WP ACF67367\_30 3000001 3110000  
 WP ACF67367\_31 3100001 3210000  
 WP ACF67367\_32 3200001 3310000  
 WP ACF67367\_33 3300001 3410000  
 WP ACF67367\_34 3400001 3510000  
 WP ACF67367\_35 3500001 3610000  
 WP ACF67367\_36 3600001 3710000  
 WP ACF67367\_37 3700001 3810000  
 WP ACF67367\_38 3800001 3910000  
 WP ACF67367\_39 3900001 4010000  
 WP ACF67367\_40 4000001 4110000  
 WP ACF67367\_41 4100001 4210000  
 WP ACF67367\_42 4200001 4310000  
 WP ACF67367\_43 4300001 4410000  
 WP ACF67367\_44 4400001 4510000  
 WP ACF67367\_45 4500001 4610000  
 WP ACF67367\_46 4600001 4710000  
 WP ACF67367\_47 4700001 4810000  
 WP ACF67367\_48 4800001 4910000  
 WP ACF67367\_49 4900001 5010000  
 WP ACF67367\_50 5000001 5110000  
 WP ACF67367\_51 5100001 5210000  
 WP ACF67367\_52 5200001 5310000  
 WP ACF67367\_53 5300001 5410000

```

WP ACF67367_54 5400001 5510000
WP ACF67367_55 5500001 5610000
WP ACF67367_56 5600001 5648894

Query Match
Best Local Similarity 61.1%; Score 278.2; DB 7; Length 110000;
Matches 544; Conservative 0; Mismatches 278; Indels 69; Gaps 3;

QY 14 ATGCGTCAGGGTATCTGACGATTTCCCAATATGACCCCGATGTTTGTGATCGGCGCC 73
DB 94457 ATGAGTAAACAGCTACCTGCGATTTCTTAATATGACCCCGATGTTTGTGATCGA 94398
QY 74 CTAGCGGTGGCTGTATGAGCTTATGATATTTGGGTTTCCCTTTTGGTATGAGTGG 133
DB 94397 ATAGCCCTCATGTTGATGATGTTTCAATGATCTTGTGTTGGTTTATTCGCAATGG 94338
QY 134 GCCAATGCGGAGCGGATCGCGGCGAGTGTGTAACGCGGTAGCAAGTCTGACTTG 193
DB 94337 GCGACACGAGGCGACGAAACCAATAGTGTGACAAACAAAGATGAGAACTTG 94278
QY 194 TTATTGCGCGGCTTTTATGATGATGATGATGATGATGATGATGATGATGATGAT 253
DB 94277 CTATTAGCAGGTTTCTGCGATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTTAT 94218
QY 254 AATTGATCTGTTCTCTGCTGACCTCTTATTTATTCAAAGTGAAGTGGCGGATG 313
DB 94217 AATTGCTGCTGCTTCTGCAACACCTTATTTGTTTAAATTTGGATGTTGGATG 94158
QY 314 TCTTTCACGCGGCGCTTATTTGGTGTGATCACCGCATGTTCTGTATGCGGTAAAC 373
DB 94157 TCTTTCACGCGTGTATGAGTGGGTTATCTGCGCAATGTTGTTGGCCGACAG 94098
QY 374 CAACGACCTCTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 433
DB 94097 AAACGCTATTTCTTACAGTTGCTGATTCATTTGCTGCTGCTGCTGCTGCTGCT 94038
QY 434 ATGGAGAGTATCGGTAATTTATGATGATGATGATGATGATGATGATGATGATGAT 490
DB 94037 ATGGAGAGATCGGTAATTTATTTATGATGATGATGATGATGATGATGATGATG 93978
QY 491 CCTTGGGCTTTTGTATTCCTA----- 512
DB 93977 CCGTGGGCGATGTTATTCACAGCTCCGTTGTAAGATATGATGCTGACAGCAT 93918
QY 513 -----ATGTCGCCCATGCGCGGCGCATCTTTCACAGCTT 547
DB 93917 CCCTCACTGTTATCACTATCTTGAAGCAATATGTTGTTCTGCAAGTCACTTCA 93858
QY 548 TATGAATTCGCTTGAAGCGGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 607
DB 93857 TATGAATTCGCTTGAAGCGGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 93798
QY 608 CCTGTCGCTAGGACGATCGGATCGGATGTTTATGATGATGATGATGATGATGATG 667
DB 93797 CCGTGGCGAATGGGAGTGTTCGTTATTTATTTATGATGATGATGATGATGATG 93738
QY 668 CTGTGGAATACGTCCTGAGACCAATGCTCAAGTGGGCTGTTTGGTGGCTTCAATT 727
DB 93737 ATAGTGAATTCCTTCCGCAACCAATGATGATGATGATGATGATGATGATGATG 93681
QY 728 ATGGGCAATCTCTCTTACCTATGATGATGATGATGATGATGATGATGATGATGAT 787
DB 93680 ATGGGCAATCTCTCTTACCTATGATGATGATGATGATGATGATGATGATGATG 93621
QY 788 TACCAACGCGGTTTATCAAGACGTTAGACGAAATATGAGTATGAGTATGAGT 838
DB 93620 TATTAACCAACGATATATGATGATGATGATGATGATGATGATGATGATGATG 93570

RESULT 13
ACF65386_0
WP Sequence split into 7 fragments LOCUS ACF65386 Accession ACF65386
WP Fragment Name Begin End

```

```

WP ACF65386_0 1 110000
WP ACF65386_1 100001 210000
WP ACF65386_2 200001 310000
WP ACF65386_3 300001 410000
WP ACF65386_4 400001 510000
WP ACF65386_5 500001 610000
WP ACF65386_6 600001 700779
ID ACF65386 standard; DNA; 700779 BP.
XX
AC ACF65386;
XX
DT 20-NOV-2003 (first entry)
XX
DE Photorhabdus luminescens nucleotide sequence #39.
XX
KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW detection; food; gene expression; plant; animal; microorganism; toxin;
KW antibiotic; bioplastic; virulence factor; disease model; plague;
KW whooping cough; gene; ds.
XX
OS Photorhabdus luminescens.
XX
PN WO200294867-A2.
XX
PD 28-NOV-2002.
XX
PF 07-FEB-2002; 2002WO-1B003040.
XX
PR 07-FEB-2001; 2001FR-00001659.
XX
PA (INSP ) INST PASTEUR.
XX
PI (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI Duchaud E, Taourit S, Glaeser P, Frangeul L, Kunst F, Danchin A;
PI Buchrieser C;
XX
DR WPI; 2003-148459/14.
XX
PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
XX useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX
PS Claim 1; SEQ ID NO 39; 1205bp; French.
XX
CC The invention relates to the isolation of genes and their encoded
CC proteins from Photorhabdus luminescens. The isolated sequences are
CC sources of probes and primers for detecting the genome of P. luminescens
CC and related species; to study polymorphisms; for gene analysis and for
CC detection/amplification of the genes. Antibodies (Ab) raised against the
CC polypeptides encoded by the genes are used for detection/identification
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC carry a gene-containing vector are used to select compounds that
CC modulate, regulate, induce or inhibit expression of the genes in plants,
CC animals or microorganisms other than P. luminescens and are able to alter
CC response or sensitivity to toxins and antibiotics produced by P.
CC luminescens. Cells transformed to express the genes are useful for
CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC bioplastic. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luminescens genes
XX
SQ Sequence 700779 BP; 190440 A; 140977 C; 164444 G; 204917 T; 0 U; 1 Other;
XX

Query Match
Best Local Similarity 61.1%; Score 278.2; DB 7; Length 110000;
Matches 544; Conservative 0; Mismatches 278; Indels 69; Gaps 3;

QY 14 ATGCTCAGGGTATCTGACGATTTCCCAATATGACCCCGATGTTTGTGATCGGCGCC 73

```

```

Db      104168 ATGAGTAAACAGCTACCTGCACTTCTATATATGACCCGGTCAATTTTTCATATAGTCCA 104227
Qy      74 CTAGCGGTGCGCTGGTATGCTTGAATGATTTGGTGGTTCTTTTGGTATGATGTTG 133
Db      104228 ATAGCCCTTCATTTGGTATGATTTCTATGTTCTTGGTGGTGGTATGATGATGATG 104287
Qy      134 GCCAATCGCCGAGCGAGTGCAGCGGAGTGGTGGACCGGTGACAAATCTCTGACTTG 193
Db      104288 GCGACACGAGGAGGACGCAAAACCAATAGTGTGGACAAAACGAAAGTAGAAGACTTG 104347
Qy      134 TTATTCGCGGCTTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 253
Db      104348 CTATATGACGATTTTGTGCGCTATTTGTTGGTGGTGGTGGTGGTGGTGGTGGTGG 104407
Qy      254 AATTTGATCTGTTCTGCTGCTGACCTCTTTATTTATTTCAAGTGGACCTGGCGGACTG 313
Db      104408 AATTTCCCTGCTTTCTGACAAACCCCTTTATTTGTTTAAAGTTTGGATGATGATGATG 104467
Qy      314 TCTTTCCACGCGGCTTATTTGGGTGTGATCACCGCAGTCTTGTATGCGGTAAGAAAC 373
Db      104468 TCTTTCCACGCGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 104527
Qy      374 CAACGACCTTTCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 433
Db      104528 AAACGCAATTTCTTCAAGTTGCTGACTTATGATGATGATGATGATGATGATGATGATG 104587
Qy      434 ATGGAGCTATCGGTATCTTATGATGATGATGATGATGATGATGATGATGATGATGATG 490
Db      104588 ATGGAGGAATCGGTATTTTATTTATTTAAACGCGAGTTGTGGGACGTGTCACTGATG 104647
Qy      491 CTTTGGGCTTTTGTATTTCCCTA-----ATGTTGCCCACTGCGCGCATCTTCAACAGTT 512
Db      104648 CCGTGGGCGATGTTATTTCCCAAGCTCCGTTGGTGAAGATATGATGATGATGATGATG 104707
Qy      513 -----ATGTTGCCCACTGCGCGCATCTTCAACAGTT 547
Db      104708 CCTCAGCTGTTATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 104767
Qy      548 TATGATATGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 607
Db      104768 TATGAAATGCGGTGGAAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 104827
Qy      608 CCTGTCGCTAGGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 667
Db      104828 CCTGCGCAATGGGAGTGTTCGCGTATTTCTTAATGATGATGATGATGATGATGATG 104887
Qy      668 CTGTTGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 727
Db      104888 ATAGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 104944
Qy      728 ATGGGCAAAATCTCTCTTATGATGATGATGATGATGATGATGATGATGATGATGATG 787
Db      104945 ATGGGCAAAATCTCTTATGATGATGATGATGATGATGATGATGATGATGATGATG 105004
Qy      788 TACAAGCGGCTTTGATCAAGACCGTGTAGACGCAAAATAGGATGATGATGATGATGATG 838
Db      105005 TATTAACACCAAGTATATAGTATCAAGAGTAAATGAACAATATCTGG 105055

RESULT 14
ACF65386_1
Continuation (2 of 7) of ACF65386 from base 100001 (Photorhabdus luminescens nucleotide
WP Sequence split into 7 fragments LOCUS ACF65386 Accession ACF65386
WP Fragment Name Begin End
WP ACF65386_0 1 110000
WP ACF65386_1 100001 210000
WP ACF65386_2 200001 310000
WP ACF65386_3 300001 410000
WP ACF65386_4 400001 510000
WP ACF65386_5 500001 610000
WP ACF65386_6 600001 700779

```

```

Best Local Similarity 61.1%; Pred. No. 4.8e-76;
Matches 544; Conservative 0; Mismatches 278; Indels 69; Gaps 3;

Qy      14 ATGCTCAGGGTATATCTGACATGTTCCCAATATTTGACCCCGATTTGTTTCATGCGCCCT 73
Db      4168 ATGAGTAAACAGCTACCTGCACTTCTATATATGATGATGATGATGATGATGATGATG 4227
Qy      74 CTAGCGGTGCGCTGGTATGCTTGAATGATTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 133
Db      4228 ATAGCCCTTCATTTGGTATGATTTCTATGTTCTTGGTGGTGGTGGTGGTGGTGGTGGTGG 4287
Qy      134 GCCAATCGCCGAGTATGCGCGGAGTGGTGGACGCTGAGCAAGTCTGACTTG 193
Db      4288 GCGACACGAGGAGCGCAAAACCAATAGTGTGGACAAAACGAAAGTAGAAGACTTG 4347
Qy      194 TTATTCGCGGCTTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 253
Db      4348 CTATATGACGATTTTGTGCGGTATTTGTGATGATGATGATGATGATGATGATGATGATGATG 4407
Qy      254 AATTTGATCTGTTCTGCTGCTGACCTCTTTATTTATTTCAAGTGGACCTGGCGGACTG 313
Db      4408 AATTTCCCTGCTTTCTGACAAACCCCTTTATTTGTTTAAAGTTTGGAGTGTGGAGT 4467
Qy      314 TCTTTCCACGCGGCTTATTTGGGTGTGATCACCGCATGTTCTGATGATGATGATGATGATG 373
Db      4468 TCTTTCCACGCGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4527
Qy      374 CAACGACCTTTCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 433
Db      4528 AAACGCAATTTCTTCAAGTTGCTGACTTATGATGATGATGATGATGATGATGATGATGATG 4587
Qy      434 ATGGAGCTATCGGTATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 490
Db      4588 ATGGAGGAATCGGTATTTTATTTATTTAAACGCGAGTTGTGGGACGTTGATGATGATGATG 4647
Qy      491 CTTTGGGCTTTTGTATTTCCCTA-----ATGTTGCCCACTGCGCGCATCTTCAACAGTT 512
Db      4648 CCGTGGGCGATGTTATTTCCCAAGCTCCGTTGGTGAAGATATGATGATGATGATGATGATG 4707
Qy      513 -----ATGTTGCCCACTGCGCGCATCTTCAACAGTT 547
Db      4708 CCTCAGCTGTTATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4767
Qy      548 TATGATATGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 607
Db      4768 TATGAAATGCGGTGGAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4827
Qy      608 CCTGTCGCTAGGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 667
Db      4828 CCTGCGCAATGGGAGTGTTCGCGTATTTCTTAATGATGATGATGATGATGATGATGATG 4887
Qy      668 CTGTTGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 727
Db      4888 ATAGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4944
Qy      728 ATGGGCAAAATCTCTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 787
Db      4945 ATGGGCAAAATCTTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5004
Qy      788 TACAAGCGGCTTTGATCAAGACCGTGTAGACGCAAAATAGGATGATGATGATGATGATG 838
Db      5005 TATTAACACCAAGTATATAGTATCAAGAGTAAATGAACAATATCTGG 5055

RESULT 15
ACF69784
ID ACF69784 standard; DNA; 876 BP.
XX ACF69784;
AC ACF69784;
DT 20-NOV-2003 (first entry)
XX
DE Photorhabdus luminescens nucleotide sequence #8251.

```

XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;  
 KW detection; food; gene expression; plant; animal; microorganism; toxin;  
 KW antibiotic; biopesticide; virulence factor; disease model; plague;  
 KW whooping cough; gene; ds.  
 XX  
 OS Photorhabdus luminescens.  
 XX  
 PN MO200294867-A2.  
 XX  
 PD 28-NOV-2002.  
 XX  
 PF 07-FEB-2002; 2002WO-1B003040.  
 XX  
 PR 07-FEB-2001; 2001FR-00001659.  
 XX  
 PA (INSP ) INST PASTEUR.  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 PI Duchaud B, Taourit S, Glaeser P, Frangeul L, Kunst F, Danchin A;  
 PI Buchrieser C;  
 XX  
 DR WPI; 2003-148459/14.  
 XX  
 PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,  
 PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.  
 XX  
 PS Claim 2; SEQ ID NO 8251; 1205bp; French.  
 XX  
 CC The invention relates to the isolation of genes and their encoded  
 CC proteins from Photorhabdus luminescens. The isolated sequences are  
 CC sources of probes and primers for detecting the genome of P. luminescens  
 CC and related species; to study polymorphisms; for gene analysis and for  
 CC detection/amplification of the genes. Antibodies (Ab) raised against the  
 CC polypeptides encoded by the genes are used for detection/identification  
 CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that  
 CC carry a gene-containing vector are used to select compounds that  
 CC modulate, regulate, induce or inhibit expression of the genes in plants,  
 CC animals or microorganisms other than P. luminescens and are able to alter  
 CC response or sensitivity to toxins and antibiotics produced by P.  
 CC luminescens. Cells transformed to express the genes are useful for  
 CC recombinant production of the proteins, particularly toxins and  
 CC antibacterials useful as insecticides, bactericides and fungicides. The  
 CC genes, proteins, vectors containing the genes and Ab are also useful  
 CC therapeutically (to treat microbial infection by bacteria or fungi that  
 CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as  
 CC biopesticides. Other uses of the genes and the proteins are as virulence  
 CC factors and for identifying targets of human diseases for which P.  
 CC luminescens is a model (particularly plague and whooping cough). This  
 CC sequence represents one of the isolated P. luminescens genes  
 CC  
 SQ Sequence 876 BP; 200 A; 177 C; 221 G; 278 T; 0 U; 0 Other;

Query Match 9.5%; Score 277.8; DB 7; Length 876;  
 Best Local Similarity 62.2%; Pred. No. 3.2e-77;  
 Matches 528; Conservative 0; Mismatches 252; Indels 69; Gaps 3;

QY 14 ATGCTCAGGGTTATGCAAGTTCCCAATATTGACCCCGATTTGTTTCGATCGCCCT 73  
 DB 1 ATGAGTAACAGCTACCTGCGCATTTCTTAATATTGACCCGTCATTTTCAATAGGTCCA 60  
 QY 74 CTAGCGGTGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 133  
 DB 61 ATAGCCCTCATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 120  
 QY 134 GCCAATCGCCGAGCGGATGCGGCGAGTGTGAGCCGTGAGCAAGTCTCTGACTTG 193  
 DB 121 GCGACACGACGAGGCGAAGCAACCATATGTTGAGCAAAAAACAAATAGAGAACTTG 180  
 QY 194 TTATTCGCGCGCTTTTAAAGTATGATGATGATGATGATGATGATGATGATGATGATG 253  
 DB 181 CTTTATGCAAGTGTGCTGCGATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 240

QY 254 AATTTGATCTGCTCTTCTGTCGACCCCTCTTATTTATTTCAAGTGTGACGTCGCGCATG 313  
 DB 241 AATTTCCCGCTCTTCTGACAAACCCCTTATTTGTTAAAGTTTGGATGTGTGGATG 300  
 QY 314 TCTTTCACGCGCGCTTATGAGTGTGATCACCGCATGTTCTGTATGCGGTAAAC 373  
 DB 301 TCTTTCACGCGGTGTTAGTGGGGGTATCTGCGCAATGTGTGTTGGCCGAGAACG 360  
 QY 374 CAAGCCACTTCTTGTGTGCGCGATTTGTTGCGCCCTTATGTCATGTTGGGG 433  
 DB 361 AAGCGCATTTCTTACAAAGTGTCTCATCTTCATGTCCTGTTAGTCCCTTGTAGGC 420  
 QY 434 ATGGACGATGCGTATCTTATGATATGATGATGATGATGATGATGATGATGATGATG 490  
 DB 421 ATGGGAGATGCGTATTTTATTTAATGCGCAGTGTGTGGGACGTGTACATTTGATACG 480  
 QY 491 CCTGCGCTTTGATTCCTTA-----ATGTTGCGCACGCGCGCATCTTACAGCTT 512  
 DB 481 CCGTGGCGATGTATTCACCAAGCTCCGTTGTGAAGATATTCATTTGCTGCAACCAT 540  
 QY 513 -----ATGTTGCGCACGCGCGCATCTTACAGCTT 547  
 DB 541 CCTCATCTGTTATTCAGTACTGAGCAATATGATGATGATGATGATGATGATGATGATG 600  
 QY 548 TATGAATTCGCTTGAAGGCGTGTCTGTTCTTATTTCTTAATTTGTTATTTGTTAA 607  
 DB 601 TATGAATTCGCTTGAAGGCGTGTCTGTTCTTATTTCTTAATTTGTTATTTGTTAA 660  
 QY 608 CCTGTCGCTAGGAGCGATTCGGAAGTGTGTTTATGCTGATACGTTACATTCGCTTC 667  
 DB 661 CCGCCCAATGCGAGATGTTCCGTTTATTTCTTAATTTGCTACGATATTTCCGAGTG 720  
 QY 668 CTTTGTGAATGCTCCGAGGACCAATGCTAGTGTGCTGTGTTGTTGCTTCA 727  
 DB 721 ATAGTCAGTCTTCCGCAACAGATGCACAACTGAGCTTATTTGACCGGATCAGT--- 777  
 QY 728 ATGGGCAAAATCTCTCTTACCTATGATGATGATGATGATGATGATGATGATGATGATG 787  
 DB 778 ATGGGCAAAATCTCTCTTATACATGATCTCGGGGGAATTTATATGATATGCGCA 837  
 QY 788 TACAAGCGC 796  
 DB 838 TATAAACAC 846

Search completed: July 31, 2004, 13:34:23  
 Job time : 770 secs

This Page Blank (250.0)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: July 31, 2004, 12:39:50 ; Search time 150 Seconds

(without alignments)  
10762.348 Million cell updates/sec

Title: US-09-700-712a-1

Perfect score: 2909

Sequence: 1 gaggaagttgtatgcagctc.....cagcgaatcggcgctgcag 2909

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA:\*

- 1: /cgn2\_6/prodata/2/ina/5A COMB. seq:\*
- 2: /cgn2\_6/prodata/2/ina/5B COMB. seq:\*
- 3: /cgn2\_6/prodata/2/ina/5A COMB. seq:\*
- 4: /cgn2\_6/prodata/2/ina/5B COMB. seq:\*
- 5: /cgn2\_6/prodata/2/ina/PCTUS COMB. seq:\*
- 6: /cgn2\_6/prodata/2/ina/backfillseq. seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	* Query Match Length	DB ID	Description
1	744.2	25.6 1830121	4 US-09-557-884-1	Sequence 1, Appl
2	744.2	25.6 1830121	4 US-09-643-990A-1	Sequence 1, Appl
3	410.8	14.1 873	4 US-09-543-681A-282	Sequence 2282, Ap
4	292.2	10.0 807	3 US-09-277-565-17	Sequence 17, Appl
5	282.4	9.7 810	4 US-09-252-991A-15221	Sequence 15221, A
6	282.4	9.7 954	4 US-09-252-991A-15232	Sequence 15232, A
7	282.4	9.1 891	4 US-09-252-991A-15227	Sequence 15227, A
8	265.8	9.1 879	4 US-09-489-039A-2365	Sequence 2365, Ap
9	208.6	7.2 834	4 US-09-328-352-3810	Sequence 2965, Ap
10	152.2	5.2 640681	4 US-09-790-988-1	Sequence 3810, Ap
11	143.8	4.9 891	4 US-09-540-336-663	Sequence 1, Appl
12	142.2	4.9 96109	4 US-09-596-002-35	Sequence 663, Appl
13	115	4.0 435	4 US-09-252-991A-15244	Sequence 15244, A
14	75.4	2.6 2288	4 US-08-553-433-1	Sequence 1, Appl
15	75.4	2.6 2288	4 US-09-553-433-1	Sequence 1, Appl
16	75.4	2.6 4137	4 US-09-499-964-2	Sequence 2, Appl
17	71.6	2.5 2058	4 US-09-252-991A-15243	Sequence 15243, A
18	64.8	2.2 1536	3 US-09-089-195-1	Sequence 29, Appl
19	64.8	2.2 1536	3 US-09-347-878-29	Sequence 1, Appl
20	64.8	2.2 1536	3 US-09-367-007C-38	Sequence 38, Appl
21	64.8	2.2 1536	4 US-09-962-665-6	Sequence 6, Appl
22	64.8	2.2 1536	4 US-09-963-333-6	Sequence 6, Appl
23	61.2	2.1 978	4 US-09-134-000C-941	Sequence 941, Appl
24	57	1.9 975	4 US-09-107-532A-2491	Sequence 2491, Ap
25	54.2	1.7 969	4 US-09-221-017B-1092	Sequence 1092, Ap
26	49	1.7 969	4 US-09-134-001C-357	Sequence 357, App

28	48.2	1.7	870	4	US-09-252-991A-15224	Sequence 15224, A
29	47.4	1.6	801	4	US-09-489-039A-2884	Sequence 2884, Ap
30	46.6	1.6	867	4	US-09-107-532A-2355	Sequence 2355, Ap
31	46.6	1.6	1014	3	US-09-230-637-1	Sequence 1, Appl
32	46.6	1.6	35100	2	US-08-770-379-17	Sequence 17, Appl
33	46.6	1.6	35100	3	US-08-757-669A-17	Sequence 17, Appl
34	46.6	1.6	35100	4	US-09-230-637-1	Sequence 17, Appl
35	45.8	1.6	5924	4	US-08-956-171E-130	Sequence 130, App
36	45.4	1.6	2335	4	US-08-961-527-275	Sequence 275, App
37	44.8	1.5	852	4	US-09-134-001C-2654	Sequence 2654, App
38	44.8	1.5	870	4	US-09-134-000C-1317	Sequence 1317, Ap
39	42.6	1.5	621	4	US-08-935-433-3	Sequence 3, Appl
40	42.6	1.5	621	4	US-09-553-132-3	Sequence 3, Appl
41	42.4	1.5	1839	4	US-09-203-895-2	Sequence 2, Appl
42	40.6	1.4	7218	1	US-08-232-463-14	Sequence 14, Appl
43	40	1.4	1568	4	US-08-936-165A-95	Sequence 95, Appl
44	39.8	1.4	4214	4	US-09-221-017B-293	Sequence 293, App
45	38.6	1.3	7218	1	US-08-232-463-14	Sequence 14, Appl

#### ALIGNMENTS

RESULT 1  
US-09-557-884-1  
Sequence 1, Application US/09557884  
Patent No. 6506581  
GENERAL INFORMATION:  
APPLICANT: Fleischmann et al.  
TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments thereof, and Uses Thereof  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: MD  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3 1/2 inch diskette  
COMPUTER: Dell Pentium  
OPERATING SYSTEM: MS DOS V6.22  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/557,884  
FILING DATE: 25-Apr-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/476,102  
FILING DATE: JUN-5-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Michelle S. Marks  
REGISTRATION NUMBER: 41,971  
REFERENCE/DOCKET NUMBER: PB186P3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 301-309-8504  
TELEFAX: 301-309-8439  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1830121 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-557-884-1  
Query Match 25.6%; Score 744.2; DB 4; Length 1830121;  
Best Local Similarity 66.7%; Pred. No. 1.7e-245;  
Matches 114; Conservative 0; Mismatches 538; Indels 27; Gaps 4;  
QY 2 AGAAGTTGTATGCTCAGGGTTATCTGACGTTCCCAATATGACCCCGATATGTTT 61

Db	958086	ATTAAGGAAATTTATGAAATTCAAATTTATTTACTTCTTCCCACTTTGATCCGATATTTT	958145
Qy	62	TCGATCGG- ----CTCTAGCGGTGGCGTGGTATGCTTGATGATTTTGGTGGTTTC	115
Db	958146	ACGCTGGGCGATAGTAAATATCGGCTTACGTTGGTATGGCTGATGATCACTTTTAAAGTTT	958205
Qy	116	CTTTTTCGATATGATGGTGGCCAAATCGCCAGCGGATATCGCGGGGAGTGTGGAGCGGT	175
Db	958206	GTTTTTGCACGTTGGCTTGGCTGGTTCGCCGTGCTAAATGCGCCAAATATGCGGTGGACAGTA	958265
Qy	176	GAGCAAGTCTCGACTCTGATCTTTATTTGCGCGGCTTTTATGGTATGATTCGATGGCCAGTT	235
Db	958266	GATCAAGTGTATAGCTTACTTTTCAACGGTTTTTATGGGGGTGTTTATGGCGGACGTGTT	958325
Qy	236	GGTATATGATATCTTCAAAATTTTGAATCTGTTCCTTGCTGACCCCTTTATTTTCAAA	295
Db	958326	GGCGATGATTTTTTTCTATATCTCGATCATTTCTTACAAAGAACCACTTATTTATTTCCCG	958385
Qy	236	GTTGTGACATCGGCGGACATGCTCTTCCACGGCGGCTTATTTGGGTGATTCACCGCANGTTC	355
Db	958386	GTTTGGGAAGGTGGATATGCTGTTCCACGGGTGGCTTATTTGGTGAATTTGTGCTATGATTT	958445
Qy	356	TGTTATGGCGGTAATAAACACGACCTCTTTTGTGTGGCGGATTTGTGTGCCCTTTA	415
Db	958446	TGGAATCTTATTTCTCAAAACGTAATTTTGGCAAAAGCATTTTGTGGCCTTTG	958505
Qy	416	GTCGATTCGTTTGGGGATGGGACGATTCGGTAACTTTATGATATGAACTTTGGGGA	475
Db	958506	ATTCGGTTGGTTTATGTTATGTTAGGCAAGATTTGGTAATTTCAATTAATCTTGAACATATGGGA	958565
Qy	476	CGAGTAAACGAGATGCTGTTGGGCTTTTGAATTCCTTAATGTGAGCC- --CACTGCCGCG	532
Db	958566	CGCGAAACCAATGTGCTTGGGCAAGATTTTCCCAATGATCTCTTTTACTGCTCGT	958625
Qy	532	CATCCTTACAGCTTTATGAATTCGCTTAAAGAGGTGTGCTGTTCTTTATCTTAAT	592
Db	958626	CATCAATCAAACTTTATGAGCTTTTAAAGGCTGTGTGTTTACGATTCGAAT	958685
Qy	592	TGGTTTATTTGTAACCTTCGTCGCTAGGACAGCGTATCCGACTGTTTTTATGCTGATAC	652
Db	958686	ATTTTATTTAAAAAACAGTCCAAATGCTGCTGTGTTGACAGTTTATCTTAATTTGGTTAT	958745
Qy	652	GGTACATTCGCTCTCTTGTAATATGTCCTGTAGCCAGATGCTCATGTTGGTCTGTTT	712
Db	958746	GGCGTCTTCGTTTATTTATGTGAATATGTGCGGAACCTGAAG- ----TGAAAAATTC	958799
Qy	712	GGTGGCTTATTTCAATGGGCAAAATCTCTCTCTTAACCTATGTGATCATCGGATATTTG	772
Db	958800	TTTGGATATTTAAACAGAGGGCAAGCCCTTTGCTTCCAGATATTTATGTGTGTCTTTC	958859
Qy	772	ATGATGTTTGGCTTTACAAAGCGCGGTTTGTATCAAGACCGTGTACAGAAAAATAGGGT	832
Db	958860	ATTATGGCTTTGGGCTTTATTCAGCG- ----AAAATGCGGTCAATAAATATGGAG	958907
Qy	832	AGTTAGTGAACAGATTTTGAATCTTTTGTACAGCGCATCGTCGATCAAGGTGTTGGGTT	892
Db	958908	ATTTTATGAAACAATATCTTGAGCTTTGTGCGCGCATTTGTAGTAAGGGAATGGGTT	958967
Qy	892	GAAATATGAACGAACGGGCAAGGTTGTTGACTGTGATTTAATGCCATTTTGAACCTACGAT	952
Db	958968	GCTAATGAACGTATCAGGTATGCAATGCTCAGCGTCATTTAATGCAATTTAGAAATATGAT	959027
Qy	952	GGGGCAACAATCAGTTTCTCTGATGATACACGCAAGATTTTGGAAAGCTGCGGTA	1012
Db	959028	GTTGGCAATTAATCAATTTCCGCTGATTTACTACCCGTAATAAGTATATGAAAGGCGGATTT	959087
Qy	1012	GCCGAGTGTCTCGGCTATATTCGTGTGTACGATATATGCGCGGATTTTGGCAATATAGGT	1072
Db	959088	GCTGAATTTTATAGTTATATGTGTGATATGCAATATGCGCGCTGATTTCCGCGCACTTGGC	959147
Qy	1072	ACCAAACTGGGATGCTATATGCAATTTTAAACAAAGCATGTGCTCAACATCTTAAACGT	1132

Db	959148	ACGAAAACCTGGGAGTGTATATGCAATGAAATAATGACAGCTTGCTTGCAATCCGACATCGT	959207
QY	1133	AAAGGTGAGATGACATGGGACCGCTGTATGSGTGTCAAGGGTAGAGCTTGGGCTAGACCT	1192
Db	959208	AGAGCGCTGTGAAGATATGGGGCGGTATATGTGTGCAAGGAGACATGGCGCTAAGCCT	959267
QY	1193	GATGTGTGTGCATATTGACCAAGTTGAAAAAAGATTGTGTGATATTGTGACCGTGGCGTTGAT	1252
Db	959268	AATGAGAAACATATGATCAGCTACGTAAATTTGTTATATCTTAACGAAAGGTATTGAT	959327
QY	1253	GAACGAGGTGAATTTCTTAACCTTCTACATCCGGGTGAATTTTCACATGGGGTGTTCGC	1312
Db	959328	GATTAAGGAGAGATTTTAACTTTTAAACCCGGGGAATTTGATCTTGTTGTCTTCG	959387
QY	1313	CCTTGCAATGTACAGCCATCATTTTTCATGTCTGGGGGATACCTTGATCTCAACAGTACT	1372
Db	959388	CCCTGTATGCAATACCCATATCTTTTTTCTCTGTGGGGGATACATTACATCTTACTACTAT	959447
QY	1373	CAGCGTTCATGTAGTGTGCGCTTGGGGTGTGAATTTCAACATGSGTGACAGTTATGTGTTG	1432
Db	959448	CAGCGTTCCTGTGAATGTTCGCTTGAGATTGAATTTCAATCAAAATTCAGGTTGTTACCTTC	959507
QY	1433	CTTGGCGTATGAGCAGATCAGAGCAAGGAAAAAGCCGGCTTGCGGTATCACAAGATCGTC	1492
Db	959508	TTAGCACTTATGTGCTCAGATCACGGGCAAAAAAGCCGGAAGCATATATATAAATTTGTG	959567
QY	1493	AATGGCACATTTTACCAAGATCACTCGAATTTGATGCGGATGTGAGCTTAACCGTGAAG	1552
Db	959568	AATGGCATATTTTATGAAAGATCAGCTTGAGTTAATGCGGAGCGTACAACTTAAACCGCAG	959627
QY	1553	CCATTTCCAGCGCCCTCAGTTCATATTCATCCAAAGATTAAACACTGAGATTGTGAA	1612
Db	959628	CCTTTCCATTTACAAATTTAGAAATTAATCCAGATATAAAAAAGCTTGAAGTTTGGAA	959687
QY	1613	ACTTGGGTCACTTGTGATTTTGACGTACACCGATATCAGTTCACAGATCTTATTCAA	1672
Db	959688	ACTTGGGTCAAGATGATGATTTTAAAGTCGTTGGCTATCAATCCAGAACCAATTAA	959747
QY	1673	TACCGGTTTTCAGTCTAAT 1691	
Db	959748	TATCCTTTTCCGCTCAAT 959766	
RESULT 2			
US-09-643-990A-1			
; Sequence 1, Application US/09643990A			
; Patent No. 6528289			
GENERAL INFORMATION:			
APPLICANT: Robert D. Fleischmann			
Mark D. Adams			
Owen White			
Hamilton O. Smith			
J. Craig Venter			
TITLE OF INVENTION: The Nucleotide sequence of			
the Haemophilus influenzae Rd Genome, Fragments			
Thereof, and Uses Thereof			
NUMBER OF SEQUENCES: 1			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: Human Genome Sciences, Inc.			
STREET: 9410 Key West Avenue			
CITY: Rockville,			
STATE: MD			
COUNTRY: USA			
ZIP: 20850			
COMPUTER READABLE FORM:			
MEDIUM TYPE: 3 1/2 inch diskette			
COMPUTER: Dell Pentium			
OPERATING SYSTEM: MS DOS v6.22			
SOFTWARE: ASCII Text			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/09/643, 990A			
FILING DATE: 23-Aug-2000			
CLASSIFICATION: <Unknown>			
/			

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/487,429  
 FILING DATE: 1995-06-07  
 APPLICATION NUMBER: 08/426,787  
 FILING DATE: 1995-04-21  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Kenley K. Hoover  
 REGISTRATION NUMBER: 40,302  
 REFERENCE/DOCKET NUMBER: PB186P1C1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 301-610-5790  
 TELEFAX: 310-309-8439  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1830121 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
 US-09-643-990A-1

Query Match 25.6%; Score 744.2; DB 4; Length 1830121;  
 Best Local Similarity 66.7%; Pred. No. 1.7e-245;  
 Matches 1134; Conservative 0; Mismatches 538; Indels 27; Gaps 4;

QY 2 AGAAGTTGTTATGCTGAGGTTATCTGACAGTTCCCAATTTGACCCCGATTGTT 61  
 DB 958086 ATNAGGAAATTTATGATTCATTTATTTACTTCTCCCACTTTGATCGAGTATTTT 958145  
 QY 62 TCGATCGC-----CCTTAGCGGTCGCTGATGAGTTGATTTGTTGCTTTC 115  
 DB 958146 ACGCTTGGGATGATGATATCGGCTTACGTTGATGCTGATGATGATGATTTT 958205  
 QY 116 CTTTGTGATGTTGTTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 175  
 DB 958206 GTTTTTCAGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTT 958265  
 QY 176 GAGCAAGTCTGATGTTATTTGCTGCTTGTGATGATGATGATGATGATGATGATG 235  
 DB 958266 GATCAAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 958325  
 QY 236 GGTATGATCTTCTACATTTGATGATGATGATGATGATGATGATGATGATGATG 295  
 DB 958326 GGGATGATTTTCTATATCTGATGATGATGATGATGATGATGATGATGATGATG 958385  
 QY 296 GTGTGATGCTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 355  
 DB 958386 GTTTGGGAGTGGATGCTGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 958445  
 QY 356 TGGTATGCTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 415  
 DB 958446 TGGATCTTATTTCTCAAAAAGCTAATTTTGGCAAAAGCTGATTTTGTGGCTT 958505  
 QY 416 GTGCCATTCGTTTGGGATGAGGATGCTGATGCTGATGCTGATGCTGATGCTGATG 475  
 DB 958506 ATTCCTTTGCTTTAGGTTAGGCAATTTGATTTCTTATATCTTGAATGAGGA 958565  
 QY 476 CGAGTAAAGGATGCTGCTTGGCTTGTATTCCTTAATGCTGCTGCTGCTGCTGCT 532  
 DB 958566 CCGGAAACGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 958625  
 QY 533 CATCTTCAAGCTTATTAATTTGCTTGAAGGCTGCTGCTGCTGCTGCTGCTGCTGCT 592  
 DB 958626 CATCAATCAACTTATTAAGCTTTTGAAGGCTGCTGCTGCTGCTGCTGCTGCT 958685  
 QY 593 TGGTATTTGTAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 652  
 DB 958686 ATTTTATTAATAAACAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 958745  
 QY 653 GGTACATTCGCTTCTTGTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 712  
 DB 958746 GCGCTCTCGCTTATTTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 958799

QY 713 GGTGCTTCATTTCAATGGGCAAAATCTCTCTTACTTATGCTGATCATCGTATTTTG 772  
 DB 958800 TTTGGATTAATTAACGAGGCAAGCCCTTCTGCTGCTGATGATGATGCTGCTTTC 958859  
 QY 773 ATGATGTTTGTCTTCAAGGCGGTTTGTATCAAGCCGTTGACGCAAAATAGGCT 832  
 DB 958860 ATTATGCTTGGCTTATTTCAAGC-----AAAAGCGCTCATTAATAATAGG 958907  
 QY 833 AGTTAGTAAACAGATTTAGATCTTTGTCAGCCATGCTGATCAAGGTTTGGCTT 892  
 DB 958908 ATTTTATGACCAATATCTTGAAGCTTGTCCCGCATTTGATGAGGAGGATAGCTT 958967  
 QY 893 GAAATGAACGAACGCGAGCGTTGTTGCTGATGATTAATGCCATTTGACTAGAT 952  
 DB 958968 GCTAATGAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 959027  
 QY 953 GTGGGCAATCAAGTTTCTTATGATGATCAACGCAAGATTTTGAAGCTGCGTA 1012  
 DB 959028 GTGGGCAATCAATTTCCGCTGATGATGATGATGATGATGATGATGATGATG 959087  
 QY 1013 GCGAGTTGCTGCGCTATTTGCTGATGATGATGATGATGATGATGATGATGATG 1072  
 DB 959088 GCTGAATTTTATGCTTATTTGCTGATGATGATGATGATGATGATGATGATGATG 959147  
 QY 1073 ACCAAACCTGGATGCTAATGCAATTTAAACAAGATGCTCAACATCTTACCGT 1132  
 DB 959148 ACGAAACCTGGATGCTAATGCAATTTAAACAAGATGCTCAACATCTTACCGT 959207  
 QY 1133 AAGGTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1192  
 DB 959208 AAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 959267  
 QY 1193 GATGATGCTAATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1252  
 DB 959268 AATGAGAAATCTATGATGATGATGATGATGATGATGATGATGATGATGATG 959327  
 QY 1253 GACGAGTGAATTTCTTAACTTCTACATCCGCGTGAATTTCAATGCGGCTTTC 1312  
 DB 959328 GATAGAGAGATTTTAACTTTTAACTTAACTTAACTTAACTTAACTTAACTT 959387  
 QY 1313 CTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1372  
 DB 959388 CTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 959447  
 QY 1373 CAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1432  
 DB 959448 CAGCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 959507  
 QY 1433 CTTGCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1492  
 DB 959508 TTAGCATTTATGCTGATGATGATGATGATGATGATGATGATGATGATGAT 959567  
 QY 1493 AATGCGCATTTTACCAAGATCAATGATGATGATGATGATGATGATGATGATG 1552  
 DB 959568 AATGCGCATTTTATGAGATGATGATGATGATGATGATGATGATGATGATG 959627  
 QY 1553 CATTCGCCAGCCCTCAAGTATCAATCAATCAATCAATCAATCAATCAATCAAT 1612  
 DB 959628 CTTTCCCATTTCAAAATTAATAATTAATAATTAATAATTAATAATTAATAAT 959687  
 QY 1613 ACTTGGGCTACTTGGATGATTTTGAACGTCACGGAATCACTTCAAGATCTTCA 1672  
 DB 959688 ACTTGGGCTCAATGATGATTTTAAAGTGTGCTATCAATCCACGAACCAATTA 959747  
 QY 1673 TACCGTTTCACTAAT 1691  
 DB 959748 TATCTTTTCCGCTAAT 959766

RESULT 3  
 US-09-543-681A-2282  
 ; Sequence 2282, Application US/09543681A  
 ; Patent No. 6605709

```

; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIORITY FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 2282
; LENGTH: 873
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-2282

Query Match      14.1%; Score 410.8; DB 4; Length 873;
Jest Local Similarity 67.0%; Pred. No. 5.2e-13;
Matches 583; Conservative 0; Mismatches 287; Indels 0; Gaps 0;

QY      820  AGCAAAATAGGGTACTAGTGAACACGATTTAGATCTTTGTACGCCATCTGTGATCA 879
DB      3    AACAAAAAGAGAGACTACATGAAGCACTATCTGGCATTTGTCTAACCCATTTACGATGA 62

QY      880  AGGCTTTGGGGTGAANAATGAACGACCGGCAAGCGTGTGTAAGTCTGATTAATGCCGA 939
DB      63  AGGTAAATGAGATGATTAATTAATTAACGACGAGAACCCGTTGTTTAACGATGAATTAATGCCGA 122

QY      940  TTTCACCTACATGATGGGCAACAATCATCTTCTCTAGTACTACACGCAAGATTTTG 999
DB      123  TCTGGAATATGATGTGTGCCAATTAACCATTTTCCACTCATTAACGCTTAAAGTTTTTA 182

QY      1000  GAAAGCTGCCGTAGCCGAGTTCCTCGCTATATTCGTGGTTACGATTAATGCGGCGATTT 1059
DB      183  CAAAGCCGCAATGGGAACTATTAGGTACTTACGTGTATGATATATGACAGACAATT 242

QY      1060  TCGCCAAATTAAGTACCAAAACCTGGGATGCTAATGCCAATTTAAACAGATGGCTCAA 1119
DB      243  TCGTGCATTTGGCTGTAAATACGTGGAAACGCAATGCTAATATAAATAGCGATGGTTAA 302

QY      1120  CAATCCTTACCGTAAAGGTGAGAGATGACATGGGACGCTGTATGGTGTTCAGGGTAGAC 1179
DB      303  TAATCCGATCGTAAAGGTGAGAGATGATATGGGCGGAGTCTATGTGTTCACAGACGCCA 362

QY      1180  TTGGGCTAAGCCGTGATGGTGTATATTTAGCACAGTTGAAAAAGATTTGTGATTTGAG 1239
DB      363  ATGGCAACGCCCTGATGGCTGCACTTTGATCAATTAAGTAAGGTAGATTAACCTTAAC 422

QY      1240  CCGTGGCTGTATGACCGAGGTGAATTTTAACTTAACTTCAACAATCCGGGTGAATTTACAT 1299
DB      423  TAAACGATTAAGACATCGTGTGAGATGATCACTTTTATATCCCGAGAAACAGCGTT 482

QY      1300  GGGGTTTGGGCCCTTGATCATGTAACGCCATCATTTTCATTTGCGGGGGATPACCTGTGA 1359
DB      483  AGGGTGTTTAGCTCATGATCATGATACATACATCTTTTTCATTTGCTCGGTGACAGGCTTTA 542

QY      1360  TCTCAACAGTACTCAAGCCTTACATGATGATGTGCCCTTGGGGTTGAAATTTCAACATGTGCA 1419
DB      543  TTTAACCTCATATCAAGATAGCTGATGTCCCATTAAGGTTGAATTTTAATCAAAATCA 602

QY      1420  GGGTTATGTGTTCTTGCGCTGATGAGCAACAGATCAACAGGAAAAACCGCGGCTGGCGTA 1479
DB      603  ATGCTTTGTATTAATTAAGCGCTGGTGGCTCAAAATCAACAGGGATTAACAGCAATT 662

QY      1480  TCACAAAGTCTGCATATGGCACTTTACCAAGATTAACCTCGAATTTGATGCGCGATGTGCA 1539
DB      663  CCATTAAGATGATTAATGCTCAACATTTATGAACCAATTAACCTCTGATGCGGATGTGCA 722

QY      1540  GCTAAAAAGTAGACATTTCCAGCGCTCAGTTCATATCAATCAAAAGATTTAAACACT 1599
DB      723  ACTAAAAAGAGAGCATTTGCCATTTACTTAAGTACATATTAATCAAAACATTTAAGACATT 782

QY      1600  GCAGGATTTTGAAACTTGGGTCCTTTGGATGATTTTGAACGTCAACCGGATATCAAGTTCCA 1659

```

Db	AGATGATATAGAAACTTGGGTACGACAGATGACTTAAAGTTAGAGTTATCAAGTCA	842
Oy	1660 CGATCTATTCAATACCGGTTTCACTCA 1689	
Db	843 CGAGCGATAAATACCGGTCACCGTGTA 872	
<p>RESULT 4</p> <p>US-09-277-565-17</p> <p>; Sequence 17, Application US/09277565</p> <p>; Patent No. 6207384</p> <p>; GENERAL INFORMATION:</p> <p>; APPLICANT: Mekalanos, John J.</p> <p>; APPLICANT: Akereley, Brian J.</p> <p>; APPLICANT: Rudin, Eric J.</p> <p>; APPLICANT: Camilli, Andrew</p> <p>; TITLE OF INVENTION: SYSTEMATIC IDENTIFICATION OF ESSENTIAL</p> <p>; TITLE OF INVENTION: GENES BY IN VITRO TRANSDUCTION MUTAGENESIS</p> <p>; FILE REFERENCE: 00742/052002</p> <p>; CURRENT APPLICATION NUMBER: US/09/277,565</p> <p>; CURRENT FILING DATE: 1999-03-26</p> <p>; EARLIER APPLICATION NUMBER: 60/079,770</p> <p>; EARLIER FILING DATE: 1998-03-27</p> <p>; NUMBER OF SEQ. ID NOS: 29</p> <p>; SOFTWARE: FastSeq for Windows Version 3.0</p> <p>; SEQ ID NO 17</p> <p>; LENGTH: 807</p> <p>; TYPE: DNA</p> <p>; ORGANISM: Haemophilus influenzae</p> <p>US-09-277-565-17</p>		
<p>Query Match 10.0%; Score 292.2; DB 3; Length 807;</p> <p>Best Local Similarity 65.0%; Pred. No. 1e-90;</p> <p>Matches 467; Conservative 0; Mismatches 243; Indels 9; Gaps 2</p>		
Oy	81 TGGCGTGTATGCGCTTGAATGTTGGGGGTTCTTTTGCATGAGTGTTGGCAATC 140	
Db	74 TAGGTTGTATGCGCTTGAATGTTGGGGGTTCTTTTGCATGAGTGTTGGCAATC 133	
Oy	141 GCCGACGCGATCGCGCGGCGAGTGGTGGACGCGTAGCAAGTCTCGACTTTATTCG 200	
Db	134 GCGGTGCTATCGCCCAATATGCGGTGGACGATGATCAAGTTGATAGCTTACTTTCA 193	
Oy	201 CCGGCTTTTAAAGTGTATGATCGGTGGCGGAGTTGGTATGATCTTCACAATTTG 260	
Db	194 ACGGTTTATGCGGCTTTTATTTGGCGGACGATGTTGGAGTATTTTCTATTAATCTCG 253	
Oy	261 ATCTGTCTCTGTGACCCCTCTTTATTTATCAAAAGTGGACATGGGGGATGCTTCC 320	
Db	254 ATATTTCTTCAAGAACCACTTATTTATTTCCGCGTTTGGGAGTGGAGATGCTTCC 313	
Oy	321 ACGGCGGCTTATTTGGGTGTGATCAACGCGATGTTCTGGTATGCGCTTAAACCAACGA 380	
Db	314 ACGGTGCTTATTTGGGTGTGATCAACGCGATGTTCTGGTATGCGCTTAAACCAACGA 373	
Oy	381 CCTTCTTTGGTGTGCGGATTTTGTGGCCCTTTAGTGCATTTGGGTTGGGATGGAC 440	
Db	374 ATTTTGGGAAACGGGTGATTTTGTGGCCCTTTAGTGCATTTGGGTTAGGCA 433	
Oy	441 GTATCGGTATCTTAAATAGTAACTTTGGGACAGTAAACGAGATGCTTGGGCTT 500	
Db	434 GAATGTGTATTTCAATTAATCTTGAACATATGGGACGCGAAACAAATGTGCTTGGCAA 493	
Oy	501 TTGTATTTCCCTATATGTTGGCC--CACTGCGCGCAATCCTTCAAGCTTATGATTCG 557	
Db	494 TGATTTTCCGATATGATCTCTTTTACGCTCTGTCATCAATCAACTTATATGACCT 553	
Oy	558 CCTTAAAGGCGGTGTTCTGTTCTTATTTCTTAAATGTTATTTGTTAACTGTCGCG 617	
Db	554 TTTTAAAGGCGGTGTTCTGTTCTTATTTCTTAAATGTTATTTTAAATAACAGTGCMA 613	
Oy	618 TAGGACGATATCGGATCTTTTATGCTGATACGATATTCGCTTCTGTGGAT 677	

Db 614 TGGCTTCTGTCAGAGTTTATCTTATTTGTTATGTCGCTTCCCTTTATTTGTTGAAT 673  
Qy 678 AGCTCCGTGAGCCAGATGCTGAGTTGGGCTGTTTGGGCTTCATTTCAATGAGGCAAA 737  
Db 674 ATGTGGGTAACTGAAG-----TTGAAAATTTCTTTGGGATTTATTAACAGAGGCAAG 727  
Qy 738 TCCCTCTCTTACTATGATCATCGGATTTGATGATGATGTTGGTCTTACAGCCG 796  
Db 728 CCTTTGCTTGGCGATGATTTATGTTGTTCTTATTTAGCTTGGGCTTATTAACGC 786

## RESULT 5

US-09-252-991A-15221  
; Sequence 15221, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 15221  
; LENGTH: 810  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15221

Query Match 9.7%; Score 282.4; DB 4; Length 810;  
Best Local Similarity 61.3%; Pred. No. 2,6e-87;  
Matches 473; Conservative 0; Mismatches 296; Indels 3; Gaps 1;

Qy 29 CTGCACTTCCCAATATGACCCCGATTTGTTTGGATGGCCCTCTAGCGGTCG 88  
Db 13 CTGACGTATCCCAATGATCATCGGTCGCTGGCCATGCGCGCTGAAGATCACTG 72  
Qy 89 TATGCTTGATGATTTGGTGGGTTTCTTTTGTATGTTGTTGCCAATGCCAGCG 148  
Db 73 TACGGGCTGATGATCTGATCGGATCGGCGGCTGCTGTCGTCGTCGCGCGAG 132  
Qy 149 GATCGCGCGGAGTGTGAGCGGTGACGCTCTGACTTGTATTCGCGGCTTT 208  
Db 133 AAGCGCTTGACCCGACCTGACCAAGAGCGCTTCCGACCTGCTTCTGCGTCGCC 192  
Qy 209 TTAGGTGATGATGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 268  
Db 193 TCGGGGTATCTCGGTGCGCGCTGCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 252  
Qy 269 CTTCGTCGACCTCTTATTTATTTCAAGTGTGACGTCGCGGATGCTTCACGCGCG 328  
Db 253 ATGCGCAACCCAGCGCTGATCTTCAGAGTCTGGAAGGCGGATGCTTCATGCGCG 312  
Qy 329 TTAGGTGATGATCAACCGCATGTTCTGATGCGCGTAAACCAACGACCTTCTT 388  
Db 313 CTGCTCGGCGTATGTCGCGGTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCG 372  
Qy 389 GGTGTCGCGATTTTGTTCCTTTCAGTTCGTCGTCGTCGTCGTCGTCGTCGTCG 448  
Db 373 CAGCTGATGATCTTATGCGCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGCG 432  
Qy 449 AACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 508  
Db 433 AACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 492  
Qy 509 CCTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 568  
Db 493 CCAACGCGCGCGCTGCGCGCGCTGCGCGCGCTGCGCGCGCTGCGCGCGCTGCG 552

Qy 569 GTGTTCTGTTCTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 628  
Db 553 GTGGGCTGTTGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 612  
Qy 629 TCCGAGCTGTTTATGCTGATGATGATGATGATGATGATGATGATGATGATGATG 688  
Db 613 TCCGCTGCTGTTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 672  
Qy 689 CCAGATGCTCAGTGGG---TCGTTTGGTGGCTTCATTTCAATGAGGCAATCTCTCC 745  
Db 673 CCGATGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 732  
Qy 746 TTAGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 797  
Db 733 GTACGATGCTGCTGCGCGGATGCTGCGGATGCTGCGGATGCTGCGGATGCTGCG 784

## RESULT 6

US-09-252-991A-15232/C  
; Sequence 15232, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 15232  
; LENGTH: 954  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15232

Query Match 9.7%; Score 282.4; DB 4; Length 954;  
Best Local Similarity 61.3%; Pred. No. 3e-87;  
Matches 473; Conservative 0; Mismatches 296; Indels 3; Gaps 1;

Qy 29 CTGCACTTCCCAATATGACCCCGATTTGTTTGGATGGCCCTCTAGCGGTCG 88  
Db 945 CTGACGTATCCCAATGATCATCGGTCGCTGGCCATGCGCGCTGAAGATCACTG 886  
Qy 89 TATGCTTGATGATTTGGTGGGTTTCTTTTGTATGTTGTTGCCAATGCCAGCG 148  
Db 885 TACGGGCTGATGATCTGATCGGATCGGCGGCTGCTGCTGCTGCTGCTGCTGCG 826  
Qy 149 GATCGCGCGGAGTGTGAGCGGTGACGCGTGAAGTCTCTGACTTGTATTCGCGCTTT 208  
Db 825 AAGCGCTTGACCCGACCTGACCAAGAGCGCTTCCGACCTGCTGCTGCGTCGCC 766  
Qy 209 TTAGGTGATGATGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 268  
Db 765 TCGGGGTATCTCGGTGCGCGCTGCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 706  
Qy 269 CTTCGTCGACCTCTTATTTATTTCAAGTGTGACGTCGCGGATGCTTCACGCGCG 328  
Db 705 ATGCGCAACCCAGCGCTGATCTTCAGAGTCTGGAAGGCGGATGCTTCATGCGCG 646  
Qy 389 GGTGTCGCGATTTTGTTCCTTTCAGTTCGTCGTCGTCGTCGTCGTCGTCGTCG 448  
Db 585 CAGCTGATGATCTTATGCGCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGCG 526  
Qy 449 AACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 508



QY 326 GGCTTATTGGTGTGATCACCAGCTTCTGGTATGCGGTAAACCAAGCACTTC 385  
 DB 319 GGCTTATTGGTGTGATCACCAGCTTCTGGTATGCGGTAAACCAAGCACTTC 378  
 QY 386 TTGGTGTGCGGATTTGTTGCTTGTAGTCCATTCGTTGGGATGAGAGATATC 445  
 DB 379 TTCAAGTGTGAGACTTGTGCTCATTAATCCATTGGTTAGTTAGAGACATAT 438  
 QY 446 GGTAATTATTAATAGTACCTTGGGAGAGATAC---GGATGCTTGGGCTTTT 502  
 DB 439 GGTAATTATTAATAGAGAGCTTGGGAGAGATACCTTAGATACCCCGGCGATT 498  
 QY 503 GATATCCC----- 510  
 DB 499 TTAATCCAGGCTCCCGTGCAGAAATCTACAGCTTGTGCGCAAGATCCACTACATA 558  
 QY 511 -----TAATGATGAGCCCACTGCGCGCATCTTCAAGCTTATAGATTC 556  
 DB 559 TTACCTATCATTCAGCAATAGCGCGATATGCTCGCCACCATCGCACTTATAGATG 618  
 QY 557 GCTTATGAAGGCGTGTGTTCTTATTTCTTAATGTTATGTTAAGTAACTGCTCG 616  
 DB 619 TTCTTAGAAGGCGTGTGTTATTTATTTCTGAATATCTTGTGCGCAAAATCGTCT 678  
 QY 617 CTAGGAGGCTATCCGAGCTGTTTATGCTGATACGGTACATTCGCTTCTGTGGA 676  
 DB 679 ATGGGAGATGATCGGCTTATTTATTTGTTATGCTTCTGATTAATGTTGA 738  
 QY 677 TACGTCGTGAGCAGATGCTGATGGTCTGTTGTTGCTTCAATTCAGAGGCA 736  
 DB 729 TTCTTCGCGCACTGATGCGCAATTAAGTTTATTAAGCGGATAC---GATGGGCA 795  
 QY 737 ATCTCTCTTACCTATGATGATCATCGTATTTGATATGTTGTTCTTACAGCC 796  
 DB 736 ATCTTTCATTCATGATGATTAATTAAGCATTAATCTTATGATGCTGATGCTCC 855  
 QY 797 GGT 799  
 DB 856 GAT 858

# RESULT 9 US-09-489-039A-2965

; Sequence 2965; Application US/09489039A  
 ; Patent No. 6610836  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary Breton et. al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
 ; FILE REFERENCE: 2709.2004001  
 ; CURRENT APPLICATION NUMBER: US/09/489,039A  
 ; PRIORITY FILING DATE: 2000-01-27  
 ; PRIOR APPLICATION NUMBER: US 60/117,747  
 ; PRIORITY FILING DATE: 1999-01-29  
 ; NUMBER OF SEQ ID NOS: 14342  
 ; SEQ ID NO 2965  
 ; LENGTH: 879  
 ; TYPE: DNA  
 ; ORGANISM: Klebsiella pneumoniae  
 US-09-489-039A-2965

Query Match 8.8%; Score 256.8; DB 4; Length 879;  
 Best Local Similarity 59.5%; Pred. No. 2.3e-78;  
 Matches 515; Conservative 0; Mismatches 282; Indels 69; Gaps 2;

QY 14 ATGGCTCAGGGTATCTGAGTTTCCATATTAACCCGTAATGTTTTCAGTCCGCT 73  
 DB 4 ATGAATAGTGGCTACCTGATTTTCGGAGTTGATCGGATCAATTTTCTCAGGCCG 63  
 QY 74 CTAGCGGTGCGGTGATGATGATGATGATGATGATGATGATGATGATGATG 133  
 DB 64 GCTCCTCCTCACTGATGATGATGATGATGATGATGATGATGATGATGATG 123

QY 134 GCCATCCCGAGCGGATGCGCGGAGTGTGAGCGCGTGAAGTCTGACTTG 193  
 DB 124 GCCATCCCGAGCGGATGCGCGGAGTGTGAGCGCGTGAAGTGAAGTGAAGTGA 183  
 QY 194 TTATTCGCGGCTTTTATGATGATGATGATGATGATGATGATGATGATGATG 253  
 DB 184 CTCTACGCGGCTTCTCGGCGTCTTCCGCGCGCGCTATCGGCTAGGCTGTTTAC 243  
 QY 254 AATTATGATCTGTTCTGATGATGATGATGATGATGATGATGATGATGATGATG 313  
 DB 244 AACCTGCGGATTTCTTGCAGATCCGCTGATGTTCCGCTGAGAGCGGCAATG 303  
 QY 314 TCTTCACAGCGGCTTATGATGATGATGATGATGATGATGATGATGATGATG 373  
 DB 304 TCGTTCACAGCGGCTTATGATGATGATGATGATGATGATGATGATGATGATG 363  
 QY 374 CAAGCAGCTTCTTGTGATGATGATGATGATGATGATGATGATGATGATGATG 433  
 DB 364 AAACGACCTTCTTCAAGTCTTCAAGTCTTCAAGTCTTCAAGTCTTCAAGTCTT 423  
 QY 424 ATGGAGCTATCGGATCTTATGATGATGATGATGATGATGATGATGATGATG 481  
 DB 424 GCGGCGCTGCGGCACTTATCAAGCGAGCTGTGCGCGGCTAGACCCGAGCTTC 483  
 QY 482 -----ACGAT 487  
 DB 484 CACTACAGATGATTTTCCCGGCTCCCGTGCAGAACCTGCGCTGCTGCCAGCAT 543  
 QY 488 GTGCTTGGGCTTTTATGATGATGATGATGATGATGATGATGATGATGATGATG 547  
 DB 544 CCGAATGCGAATGCTGTTGATGATGATGATGATGATGATGATGATGATGATG 603  
 QY 548 TATGATTCGCTTGAAGGCGTGTGTTCTTATTTCTTATTTCTTATTTCTTAT 607  
 DB 604 TACGAGCTGCGGCTGAAGGCGTGTGATGATGATGATGATGATGATGATGAT 663  
 QY 608 CCGTGTGCGGCTATGAGGAGGATTCGCGGCTGTTTATGCTGATGATGATGATG 667  
 DB 664 CCGCGCCGAGCGGCTCGTGTGCTGTTGCTGATGATGATGATGATGATGATG 723  
 QY 668 CTGTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 724  
 DB 724 ATCGTGAATTTCTTCCGACGAGCGGATGATGATGATGATGATGATGATGAT 783  
 QY 725 TCAATGCGGCAATCTTCTTATGATGATGATGATGATGATGATGATGATGATG 784  
 DB 784 AGCATGCGGAGATTTCTTCAATCCGATGATGATGATGATGATGATGATGAT 843  
 QY 785 TCTTACAGCGGCTTGTATCAAGA 810  
 DB 844 GCTTACCGGCAACCGTCCGAGCCACA 869

## RESULT 10

US-09-328-352-3810  
 ; Sequence 3810; Application US/09328352  
 ; Patent No. 6562958  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary L. Breton et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
 ; FILE REFERENCE: GTC99-039A  
 ; CURRENT APPLICATION NUMBER: US/09/328,352  
 ; PRIORITY FILING DATE: 1999-06-04  
 ; NUMBER OF SEQ ID NOS: 8252  
 ; SEQ ID NO 3810  
 ; LENGTH: 834  
 ; TYPE: DNA  
 ; ORGANISM: Acinetobacter baumannii  
 US-09-328-352-3810

Query Match 7.2%; Score 208.6; DB 4; Length 834;

Best Local Similarity 58.0%; Pred. No. 1.4e-61;  
Matches 452; Conservative 0; Mismatches 309; Indels 18; Gaps 4;

```
OY 29 CTGCAGTTTCCCAATATTGACCCCGATTTGTTGATCGAGCCCTCTAGCGGCTGCGCTG 88
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 19 CTGACCTATCTCAATATGATCGAGTCCGATCAATCAATCTAGGACCTCTTCAAGTCCATTGG 78
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 89 TATGCGTTGATGATTTGGTGGGTTTCTTTTGTATGTTGGTGGCCATTCGCCAGCG 148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 79 TATGACATCATATTTATTTGACATTTTATGCTTGGGGCTTGGCTTCTTCAAGTGGCC 138
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 149 GATCGCGGGGAGATGTTGACCGGTGAGCAAGTCTCTGACTTGTATGTCGGCGCTT 208
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 139 AAACG---GCGTGAATGCTGACATCGACATGATGTTCCGATCTGTGTTTACGCGCC 195
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 209 TTAGGTAGTATGATCGGTGCGGAGTGGTATGATGATCTTACATTTTATGATCTGTTTC 268
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 196 TTAGGTGTTGTAATCGGCGGAGCTATCGCTATGTCCTTTTCTATGAGTTGATTAATTC 255
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 269 CTTCGTGACCCCTCTTATTTATTTCAAGTGTGACCTGGCGGATGCTTCCACGCGGCG 328
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 256 CTGGAACCCCAATCTGTTATTTCCAGTCTGACAGGTGGATAGAGCTTCCACGCGGCG 315
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 329 TTATGCGGTGATCAACGCGCATGTTCTGTATGCGCGTAAACCAACGACCTTCTTT 388
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 316 TTCTTAGGGGTGATGATCGCAATGCTATCTGTGTAATAAATAATCAAAAAACATGCTTC 375
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 389 GGTGTGCGCATTTTGTGCCCCCTTTAGTGCATTCGTTGGGGATGGGAGCTATCGGT 448
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 376 CAGACATCTGACTTGTGTCCTTGTGTGTAACGACAGGCTTATATGTTGAGAGTATGGT 435
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 449 AACTTATGATATGTAATCTTTGAGGAGCG-----AGTAACGAGATGCTTGGGCT 499
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 436 AACTTATTTGGGAGAGATATATGTTGTTGGGTCAGACAGATCCGAATATTCATTCCGA 495
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 500 TTTGATTTCCCTA---ATGTTGGCCACTGCGCGGCATCTTCAACAGTTTATGAATTC 556
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 496 ATGATTTTCCCAACGAGATCAATTAATCTGTTGCTGTCACCTTCAATCAATCAACAGCG 555
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 557 GCTTAGAAGCGGTGTTCTGTTCTTTATCTTAAATGTTTATGTAATCTGTCG 616
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 556 CTTCGGAAGGACTTACTCTTTATTTTGTGTTGTTAGTCAAAAACAGCCCA 615
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 617 CTAGCAGCGATTCGCGACTGTTTATGCTGATACGTAATCCGCTTCTTGTGAA 676
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 616 CCTATGCGCTATCTGCTTATTTTAAAGGTTAAAGGTTGACAGCTTTGTATGAA 675
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 677 TACGTCGTGACGATGCTCAGTTGGT---CTGTTGTCCTTCAATTTCAATGGCG 733
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 676 TTCTTCCGTCAGCCGAGATGCCAGCAAGGTTTCATCTTGTGCTGATGACCAAGCA 735
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 734 CAATCTCTCTTACTATGATGATCAATCGGTATTTGATGATGATTTGCTTCAAA 792
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 736 CAATCTCTCAGCTTCAATGCTACTGATGAGCTTTGATGATGATGATGCTTACCA 794
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

## RESULT 11

```
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: MATANABE, HIDEKI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
```

; SEQ ID NO 1  
; LENGTH: 640681  
; TYPE: DNA  
; ORGANISM: Buchnera sp.  
US-09-790-988-1

Query Match 5.2%; Score 152.2; DB 4; Length 640681;  
Best Local Similarity 52.6%; Pred. No. 1.2e-39;  
Matches 432; Conservative 0; Mismatches 323; Indels 66; Gaps 2;

```
OY 35 TTTCCAATATTGACCCCGATGTTTGCATTCGAGCCCTCTAGCGGCTGCGGTATGCG 94
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 479304 TTCTCAATTAATTAATCTCAATATTTTACATGTCGCTGATCTGCTCGCTGATAGGT 479363
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 95 TTGATGATTTGTTGGGTTTCTTTTGTATGTTGTTGGCAATCCGCGAGCGATTCG 154
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 479364 TTATGATATGATTAATAGTTTATTTATGCTATGTTGATGAAAAAAAGTGCAGATTAA 479423
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 155 GCGGCGAGTGTGAGCGCGTGAAGACCTCTGACTTGTATTCGCCGCTTTTAGGT 214
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 479424 AATAAAAAAATATGTAAGAAAAAATAAGAAACATTAATTAATTTATTTTAAAGGA 479483
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 215 GTAATGATCGGTGCGGAGTGTGATGATCTTCTCAATTTTGAATCTGTCTTCT 274
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 479484 TCATGATCGGAGAGAAATGATATATATTTTATTAATTTTCAATTAATTTCAA 479543
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 275 GACCCCTTTTATTTCAAGTGTGACCTGCGGCAATGTCCTTCCACGCGGCTTATTTG 334
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 479544 AATATGTTATCTGATTTTATATATGGAAGAGAGCATGATTCATGAGAGATTATA 479603
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 335 GGTGTGATCACCGCATGTTCTGTATGCGGTAAACCAACGACCTTCTTGTGTG 394
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 479604 GGTGTATATATGTCATGTCATATTTTCTTTAAATTAATAAAAAAATTTAGAAATA 479663
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 395 GCCATTTTGTGCCCTTTGTCGCAATGCGTTTGGGAGATGGAGCATGATGTAATCTT 454
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 479664 TGTATTTTCACTCTCACTAATACCTTTGTTAGCGCTGTGATGAATGAAACTTT 479723
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 455 ATGAATAGTGAATCTTGGGAGAGAGT---AACGATGTCCTTGGGCTTTGTATTCCT 511
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 479724 ATTATATGTAATATAGGGTGTGTATCACCCAAATTTTCAATATGCAATATTTTCC 479783
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 512 AATGCT----- 517
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 479784 AATCTCAAAATCAAGATTTAAAGAAATAAAAAATATCCGATTAACAATTAATTA 479843
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 518 -----GCCCACTGCCCGCCATCCCTTCAACAGCTTATGATTCGCTTGAAGGC 568
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 479844 GATCAATACGAGGATTTGCGGTCATCTTACCAATTAATGAAATTTTGTAGAGGG 479903
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 569 GTGGTTCGTTCTTATTTCTAATTTGTTTATTTGTTAACTCTGCGCTAGGACGCTA 628
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 479904 ATTCTTCTGTTTATTAATTTATTTTTCGAAAAAGATACCAACAGATAGTAT 479963
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 629 TCCGACGTTTTTAACTGATACGTAATTCGCTTCTTGTGAATACGTCGCTGAG 688
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 479964 AGCGTTATTTTAAATTTTATGATTAATTAAGAAATTAATTAAGAAATTTTAAAGAA 480023
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 689 CCAATGCTCAGTGGGCTGTGTTGTTGAGCTTCAATTTCAATGAGGCAAAATCTCTCTTA 748
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 480024 CCAATTCACAAATAGGACTTTTAAATAATTAATTTCTATGAGGACAAATATTTACACTT 480083
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 749 CCTATGATCATCGATTTTGTATGATGTTGTTGCTTGA 789
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 480084 CCTATGATTTGCTGATTAATTAATTAATTAATTTTA 480124
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

## RESULT 12

```
US-09-540-236-663
; Sequence 663, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
```

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS  
FILE REFERENCE: 2709.2005-001  
CURRENT APPLICATION NUMBER: US/09/540,236  
CURRENT FILING DATE: 2000-04-04  
NUMBER OF SEQ ID NOS: 3840  
LENGTH: 891  
TYPE: DNA  
ORGANISM: M. catarrhalis  
US-09-540-236-663

Query Match  
Best Local Similarity 59.4%; Score 143.8; DB 4; Length 891;  
Matches 263; Conservative 0; Mismatches 177; Indels 3; Gaps 1;

QY 38 CCCAATATTGACCCCGTATTGTTTTCGATCGCCCTTACGCGGTGCGGTGATGCTTG 97  
DB 28 CCACAGATTAACTCTGCTTGTAGACTTGGCATATGAGCTGATGATGCTTG 87  
QY 98 ATGATTTGGTGGTTCCTTTTGTCTATGTGTTGGCCATGCGGAGCGATCGCGG 157  
DB 88 ATGATCTTATTTGCTTTTGTGGTGGCTTATTTTGGCATGCGGTGCAACAA--ATCC 144  
QY 158 GGCAGTGTGAGCGCGTGAAGAGTCTGACTTGTATTCGCCGCTTTTATAGGTGA 217  
DB 145 CGAAGCATTTTAACTCTGATGATGTTGGATTTGTTTGTTCAGTGGGTG 204  
QY 218 GTGATCGTGGCGGATGTTGTTGATGATCTTCTACAAATTTGATCTGTTCTGCTGAC 277  
DB 205 ATTTTGTGGCGGATTTGCTGATGATGATTTTATACAAATTTGATGATGATGCTAAC 264  
QY 278 CCTCTTATTATTAACAAGTGTGAGTGTGCGGATGCTCTTCCAGCGCGCTTATGGGT 337  
DB 265 CCTTATATCTTTTTCAGATGATGGAAGGCGGATGATGTTTCTGATTTGTTG 324  
QY 338 GTGATACCGCATGTTCTGTATGCGCGTAAACCAACGACCTTCTTGGTGTGCGC 397  
DB 325 GTGTTGCTTCCATGATGATTTTTCACATTAATTAATAAAGACACCTTTGGGTACG 384  
QY 398 GATTTTGTGCCCCCTTATGATGCTTGGTGGGATGGAAGTATCGTAACTTTATG 457  
DB 385 GATTTATGCAACCTGTTGACCAACAGGCTTATTTTGGGGAATTTGGAATTTTATC 444  
QY 458 AATAGTGAACCTTTGGGAGAGT 480  
DB 445 AATGGAGCTTTGGGAGAGAT 467

RESULT 13  
US-09-596-002-35  
Sequence 35, Application US/09596002  
Patent No. 6632636  
GENERAL INFORMATION:  
APPLICANT: Lagace, Robert, E.  
APPLICANT: Paterson, Chandra  
APPLICANT: Berg, Kim, L.  
TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME  
FILE REFERENCE: PM-0008-4 US  
CURRENT APPLICATION NUMBER: US/09/596,002  
CURRENT FILING DATE: 2000-06-16  
PRIOR APPLICATION NUMBER: 60/140,121  
PRIOR FILING DATE: 1999-06-18  
NUMBER OF SEQ ID NOS: 41  
SOFTWARE: PERL Program  
SEQ ID NO 35  
LENGTH: 96109  
TYPE: DNA  
ORGANISM: M. catarrhalis  
FEATURE:  
NAME/KEY: misc. feature  
OTHER INFORMATION: Incyte template ID No. 6632636 35  
PUBLICATION INFORMATION:

US-09-596-002-35

Query Match  
Best Local Similarity 59.1%; Score 142.2; DB 4; Length 96109;  
Matches 262; Conservative 0; Mismatches 178; Indels 3; Gaps 1;

QY 38 CCCAATATTGACCCCGTATTGTTTTCGATCGCCCTTACGCGGTGCGGTGATGCTTG 97  
DB 51130 CCACAGATTAACTCTGCTTGTAGACTTGGCATATGAGCTGATGATGCTTG 51189  
QY 98 ATGATTTGGTGGTTCCTTTTGTCTATGTGTTGGCCATGCGGAGCGATCGCGG 157  
DB 51190 ATGATCTTATTTGCTTTTGTGGTGGCTTATTTTGGCATGCGGTGCAACAA--ATCC 51246  
QY 158 GGCAGTGTGAGCGCGTGAAGAGTCTGACTTGTATTTGCGGCTTTTATAGGTGA 217  
DB 51247 CGAAGATTATTAACCTCTGAGATGTTTGGATTTTGTGTTTGTTCAGTGGGTG 51306  
QY 218 GTGATCGTGGCGGATGTTGATGATCTTCAAAATTTGATCTGTTCTGCTGAC 277  
DB 51307 ATTTAGTGGCGGATTTGCTATGATGATCTTATACAAATTTGATGATGCTAAC 51366  
QY 278 CCTCTTATTATTAACAAGTGTGACTGCGGATGCTCTTCCAGCGCGCTTATGGGT 337  
DB 51367 CCTTATATCTTTTTCAGATGATGGAAGGCGCATGAGTTTTCATGATGATTTGTTG 51426  
QY 338 GTGATACCGCATGTTCTGTATGCGGTAAACCAACGACCTTCTTGTGTGCGC 397  
DB 51427 GTGTTGCTTGCATGATGATTTTGCATTAATTAATAAAGACACCTTTACGTTACG 51486  
QY 398 GATTTTGTGCCCCCTTATGATGCTTGGTGGGATGGAAGTATCGTAACTTTATG 457  
DB 51487 GATTTATGCAACCTGTTGACCAACAGGCTTATTTTGGGGAATTTGGAATTTTATC 51546  
QY 458 AATAGTGAACCTTTGGGAGAGT 480  
DB 51547 AATGGAGCTTTGGGAGAGAT 51569

RESULT 14  
US-09-252-991A-15244/C  
Sequence 15244, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rudenfeld et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 15244  
LENGTH: 435  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15244

Query Match  
Best Local Similarity 59.2%; Score 115; DB 4; Length 435;  
Matches 196; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 29 CTGAGTTTCCCAATATTGACCCCGTATTGTTTTCGATCGGCGCTCTAGCGGTGCGCTG 88  
DB 373 CTGACGTATCCCAATGATGATCGTTGCGTGGCATATGAGCTGATGATGCTG 314  
QY 89 TATGCTTGTATTTGTTGTTGTTTCTTTTGTCTATGTTGTTGTTGTTGTTGTTGTTG 148  
DB 313 TACGGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 254

QY 149 GATGCGGGGAGTGGTTGACGCGTGAAGTCTCTGACTTGTATTGCGCGCTTT 208  
DB 253 AAGCGCTTGAACCCGACCGTGAAGAGCGCTTTCGACCTGCTTCTGCGGTCCGC 194  
QY 209 TTAGGTGAGTATGCGTGGCCGAGTGTATGATGATCTTAAATTTGATCTGTT 268  
DB 193 TCGCGGTGATCTCGTGGCCGCTGCGTACGTCGTTCTTACAACTGAGCACTAC 134  
QY 269 CTTCGACCCCTCTTATTATTATCAAAAGTGTGACTGCGCGAGTCCCTTCCACGCGGC 328  
DB 133 ATCGCAACCCGACCGCTATCTTCGAGGCTGTGAAGAGGCGCATGCTTCATGCGCGC 74  
QY 329 TTATTGGGTGATCAACGCGCATGTTCTGCT 359  
DB 73 CTGCTGCGGTGATGCTGCGGTCTGCTGCT 43

## RESULT 15

US-08-935-433-1/c  
; Sequence 1, Application US/08935433  
; Patent No. 6319688  
; GENERAL INFORMATION:  
; APPLICANT: FEILD, JOHN  
; TITLE OF INVENTION: A HUMAN SODIUM DEPENDENT PHOSPHATE  
; TITLE OF INVENTION: TRANSPORTER (1PT-1)  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: RATNER & PRESTIA  
; STREET: P.O. BOX 980  
; CITY: VALLEY FORGE  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19482  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/935,433  
; FILING DATE: 23-SEP-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/044,974  
; FILING DATE: 28-APR-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PRESTIA, PAUL F  
; REGISTRATION NUMBER: 23,031  
; REFERENCE/DOCKET NUMBER: GH-70006  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-407-0700  
; TELEFAX: 610-407-0701  
; TELEX: 846169  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2288 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; US-08-935-433-1

Query Match 2.6%; Score 75.4; DB 4; Length 2288;  
Best Local Similarity 56.0%; Pred. No. 7.8e-15;  
Matches 164; Conservative 0; Mismatches 126; Indels 3; Gaps 1;

QY 2476 CATTTCTAGAGGTAGAGATGACACCGCAGAAAGATTGAAAAATCGTGGATGAGTGC 2535  
DB 771 CACCTTCACGGGCAAGAGACCAACGACGACGACGACGACGACGACGACGACGACGACG 712  
QY 2536 ACTGCGAAAGACGCGGCAAACTCTTCTTTACAGCGCATATGCGCAAGGCTGACGAGAGT 2595  
DB 711 TCTCTCAAAAGCTTTCTGAACCTCCGATCTCCACCTGATAGCGCAACAATAGT 652

QY 2596 ATTGCTCACTAGTATACCAATATTGGACCCCATCACTAAGAAATCGCGTTTCAACCG- 2654  
DB 651 GTTGGTATGACGTTCCAAATGTTGGCCCCCATGATATATGGGATGCGACGCCCAACAGT 592  
QY 2655 --GTAAACCAACGGCAAGAGACCAACAAATATGAAGTCAACGCTGTTGAGATTGAAT 2712  
DB 591 GAGCAATGAAAGAGACACATGCTGACCAACGATGAGCGTTGAGGTGCTGAGCTCTGCAC 532  
QY 2713 CAGTCCGTTGCCACTAAACCAATCATCATCTGCAATTGGGTGGAGCA 2765  
DB 531 CAAAGCGTCAACAGACCCCGATCACAGGCCCAAGAGGTTGACATTA 479

Search completed: July 31, 2004, 17:02:39  
Job time : 160 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2004, 15:39:05 ; Search time 888 Seconds  
(without alignments)  
16062.201 Million cell updates/sec

Title: US-09-700-712a-1

Perfect score: 2909

Sequence: 1 gagaagcttgcattgcctc.....cagcgaatcgcgctgcag 2909

Scoring table: IDENTITY\_NUC  
Gapop: 10.0, Gapext: 1.0

Searched: 3222919 seqs, 2451570024 residues

Total number of hits satisfying chosen parameters: 6445838

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Published Applications NA.\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq2:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1112.6	38.2	1149	13	US-10-282-122A-40937
2	842.4	29.0	852	13	US-10-282-122A-40936
3	744.2	25.6	1830121	15	US-10-339-960-1
4	744.2	25.6	1830121	16	US-10-339-960-1
5	505.4	17.4	852	13	US-10-282-122A-30692
6	462.2	15.9	852	13	US-10-282-122A-22162
7	403.8	13.9	849	13	US-10-282-122A-12335
8	248.2	8.5	876	9	US-09-741-669-182
9	152.2	5.2	640681	9	US-09-790-988-1
10	142.2	4.9	96109	13	US-10-672-787-35
11	115.4	4.0	2731748	17	US-10-297-465A-1
12	84	2.9	2440	12	US-10-152-319A-1589
13	77.8	2.7	3950	12	US-10-152-319A-1988
14	75.4	2.6	2208	13	US-10-276-774-1331

15	75.4	2.6	2280	16	US-10-173-999-149	Sequence 149, App
16	75.4	2.6	3952	13	US-10-116-802-36	Sequence 36, Appl
17	75.4	2.6	4137	14	US-10-052-664-2	Sequence 2, Appl
18	73.8	2.5	4111	13	US-10-257-121-146	Sequence 146, App
19	73.8	2.5	4135	15	US-10-097-320-220	Sequence 220, App
20	72.2	2.5	687	17	US-10-404-460-258	Sequence 258, App
21	70.2	2.4	1800	13	US-10-169-395-71	Sequence 71, Appl
22	70.2	2.4	2016	13	US-10-169-395-81	Sequence 81, Appl
23	70.2	2.4	2124	16	US-10-108-260A-1011	Sequence 1011, Ap
24	70	2.4	1164	9	US-09-738-626-3017	Sequence 3017, Ap
25	70	2.4	1333	13	US-10-627-476-511	Sequence 511, App
26	70	2.4	3309400	9	US-09-738-626-1	Sequence 1, Appl
27	69.2	2.4	969	13	US-10-282-122A-33224	Sequence 33224, A
28	68.2	2.3	795	13	US-10-282-122A-42178	Sequence 14737, A
29	67.6	2.3	969	13	US-10-282-122A-14737	Sequence 14737, A
30	66.4	2.3	792	13	US-10-282-122A-12730	Sequence 12730, A
31	65.4	2.2	722	17	US-10-332-859-254	Sequence 254, App
32	64.8	2.2	1478	15	US-10-084-817-332	Sequence 332, App
33	64.8	2.2	1536	11	US-09-877-178-11	Sequence 11, Appl
34	64.8	2.2	1536	13	US-10-342-887-556	Sequence 556, App
35	64.8	2.2	1536	13	US-10-172-118-556	Sequence 556, App
36	64.8	2.2	1536	13	US-10-388-360-350	Sequence 350, App
37	64.8	2.2	1536	16	US-10-159-563-244	Sequence 244, App
38	64.8	2.2	1536	17	US-10-283-975A-254	Sequence 254, App
39	64.8	2.2	3298	15	US-10-240-965-200	Sequence 200, App
40	64.6	2.2	795	9	US-09-741-669-181	Sequence 181, App
41	64.6	2.2	795	13	US-10-282-122A-6574	Sequence 6574, Ap
42	64.6	2.2	798	9	US-09-738-626-942	Sequence 942, App
43	64.6	2.2	1200	9	US-09-954-197-1	Sequence 1, Appl
44	64.6	2.2	2730	9	US-09-987-446-1	Sequence 1, Appl
45	64.4	2.2	969	13	US-10-282-122A-14234	Sequence 14234, A

#### ALIGNMENTS

RESULT 1  
US-10-282-122A-40937/c  
Sequence 40937, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl  
APPLICANT: Zykwind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22

```
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40937
; LENGTH: 1149
; TYPE: DNA
; ORGANISM: Vibrio cholerae
US-10-282-122A-40937
```

```
Query Match      38.2%; Score 1112.6; DB 13; Length 1149;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 1118; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
```

```
QY 1783 TTAAAGCGGTGAGGCAAGATGCTGCGGGGATGACGACAAACACCAATTAAGTACT 1842
|
|
|
Db 1149 TTAAAGCGGTGAGGCAAGATGCTGCGGGGATGACGACAAACACCAATTAAGTACT 1090
|
|
|
QY 1843 CACCAACCATTTGCTCTTAACAAGCCCAAGTTGATGAGTGCACCTTTAATAGG 1902
|
|
|
Db 1089 CACCAACCATTTGCTCTTAACAAGCCCAAGTTGATGAGTGCACCTTTAATAGG 1030
|
|
|
QY 1903 CAGTTGCGCTAAGAAAGAAATACCGTAATCAAGACCGTAGCCATCAAGTTAAAGCTTAA 1962
|
|
|
Db 1029 CAGTTGCGCTAAGAAAGAAATACCGTAATCAAGACCGTAGCCATCAAGTTAAAGCTTAA 970
|
|
|
QY 1963 GTGACACGCGCAATTTGCAAGACAAACACGCGCAATCTCAACAGACACGCGGTGGCGG 2022
|
|
|
Db 969 GTGACACGCGCAATTTGCAAGACAAACACGCGCAATCTCAACAGACACGCGGTGGCGG 910
|
|
|
QY 2023 GAGCAGAGAGTAATCAAGTCCCAATGTCGACCTAAGTAATGGTGAATTTTCACG 2082
|
|
|
Db 909 GAGCAGAGAGTAATCAAGTCCCAATGTCGACCTAAGTAATGGTGAATTTTCACG 850
|
|
|
QY 2083 CACTTTAGCAGACGCGCAAGCCCAAGAGAAACCAATTAGCTGTGTGTGTCATGAAAGA 2142
|
|
|
Db 849 CACTTTAGCAGACGCGCAAGCCCAAGAGAAACCAATTAGCTGTGTGTGTCATGAAAGA 790
|
|
|
QY 2143 TTGAATTAATACCGTAACCACTGTACGTAAGCAATACCGTGTAGTGGCGCTTGGCCAT 2202
|
|
|
Db 789 TTGAATTAATACCGTAACCACTGTACGTAAGCAATACCGTGTAGTGGCGCTTGGCCAT 730
|
|
|
QY 2203 GCATTTTGTAGAAATTTCAAGTGGCGGCAACCAATCAACTCTTCATAGTTGGCCAT 2262
|
|
|
Db 729 GCATTTTGTAGAAATTTCAAGTGGCGGCAACCAATCAACTCTTCATAGTTGGCCAT 670
|
|
|
QY 2263 CACCGTAATGCGACGCAAAATGCTGCAATACCAATAGTGAAGTGCACACCGAA 2322
|
|
|
Db 669 CACCGTAATGCGACGCAAAATGCTGCAATACCAATAGTGAAGTGCACACCGAA 610
|
|
|
QY 2323 AGTATTAATCCCAATACCGAAAGCTGGGTTCAAGCCCTGTGATGACAGTTTGGTAATCGG 2382
|
|
|
Db 609 AGTATTAATCCCAATACCGAAAGCTGGGTTCAAGCCCTGTGATGACAGTTTGGTAATCGG 550
|
|
|
QY 2383 TTTGATTAATCAAAACCTTTATGCTCATATGCGCAGTGCAGAGGAGGCGAAACGAG 2442
|
|
|
Db 549 TTTGATTAATCAAAACCTTTATGCTCATATGCGCAGTGCAGAGGAGGCGAAACGAG 490
|
|
|
QY 2443 CCAAGTGTGAGACTTTCTCTTAATATGCAAAATCAATTTTGAAGGTAGAGAAATGACGAC 2502
|
|
|
Db 489 CCAAGTGTGAGACTTTCTCTTAATATGCAAAATCAATTTTGAAGGTAGAGAAATGACGAC 430
|
|
|
QY 2503 CGCGAGAGATTTGAAAAATCGTGATGTGCGACTGCGGAAAGCAACGCGGAAACTCTTC 2562
|
|
|
Db 429 CGCGAGAGATTTGAAAAATCGTGATGTGCGACTGCGGAAAGCAACGCGGAAACTCTTC 370
|
|
|
QY 2563 TTTAAGCGCGCATATGCGCAAGGCTGACGAGATATTTGTCACAGATATTCATATTTGGC 2622
|
|
|
Db 369 TTTAAGCGCGCATATGCGCAAGGCTGACGAGATATTTGTCACAGATATTCATATTTGGC 310
|
|
|
```

```
QY 2623 ACCCATCAACATAGAGATGCGGGTTTCAACCGGTAAACCCAGCGCAAGACCAACAT 2682
|
|
|
Db 309 ACCCATCAACATAGAGATGCGGGTTTCAACCGGTAAACCCAGCGCAAGACCAACAT 250
|
|
|
QY 2683 AATAGAACTCACCGTCTTGAAGATTGAATGAGCGCGTGGCACTAAACCAATCATCA 2742
|
|
|
Db 249 AATAGAACTCACCGTCTTGAAGATTGAATGAGCGCGTGGCACTAAACCAATCATCA 190
|
|
|
QY 2743 TCTGCAATTTGGGTGGAGCAAAATTCAAATAGAACTTTGGCTTGATGCGCGTTGCCA 2802
|
|
|
Db 189 TCTGCAATTTGGGTGGAGCAAAATTCAAATAGAACTTTGGCTTGATGCGCGTTGCCA 130
|
|
|
QY 2803 TTTAAACCGCTGCCGCAACATCGGACTGCAAGAAATGTAATACAGATGAAGCCAA 2862
|
|
|
Db 129 TTTAAACCGCTGCCGCAACATCGGACTGCAAGAAATGTAATACAGATGAAGCCAA 70
|
|
|
QY 2863 GTTGGCCCAAGGTAGGCTTTGTTGTGAGGAAATGCGCGCTGAG 2909
|
|
|
Db 69 GTTGGCCCAAGGTAGGCTTTGTTGTGAGGAAATGCGCGCTGAG 23
|
|
|
```

## RESULT 2

US-10-282-122A-40936

Sequence 40936, Application US/10282122A

Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu

APPLICANT: Zamudio, Carlos

APPLICANT: Malone, Cheryl

APPLICANT: Haeselbeck, Robert

APPLICANT: Ohlsen, Karl

APPLICANT: Zyskind, Judith

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John

APPLICANT: Carr, Grant

APPLICANT: Yamamoto, Robert

APPLICANT: Forsyth, R.

APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO 40936

LENGTH: 852

TYPE: DNA

ORGANISM: Vibrio cholerae

US-10-282-122A-40936

Query Match

Best Local Similarity 99.3%; Pred. No. 1.2e-267;

29.0%; Score 842.4; DB 13; Length 852;

```
Matches 846; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 839 GTGAAACAGTATTAGATCTTTGTGAGCCGATGTCATCAAGTGTGGTTGAAAT 898
| | | | |
Db 1 GTGAGCATTATTATATCTTTGTGAGCCGATGTCATCAAGTGTGGTTGAAAT 60
QY 899 GAAACGACGGGCAAGCGTGTGTGACTGTGATTAATGCCGATTGACCTACGATGTGGC 958
| | | | |
Db 61 GAAACGACGGGCAAGCGTGTGTGACTGTGATTAATGCCGATTGACCTACGATGTGGC 120
QY 959 AACCATCATTTTCTCTAGTACTACACCGAAGAGTTTGGAAACCTGCCGACGAG 1018
| | | | |
Db 121 AACCATCATTTTCTCTAGTACTACACCGAAGAGTTTGGAAACCTGCCGACGAG 180
QY 1019 TTGCTCGGCTATATTCGTGGTTCAGTAATAGCGCGGATTTTGGCAATTAAGTACCAA 1078
| | | | |
Db 181 TTGCTCGGCTATATTCGTGGTTCAGTAATAGCGCGGATTTTGGCAATTAAGTACCAA 240
QY 1079 ACCTGGGATGCTAATGCCAATTTAAACCAAGCATGGCTCAACATCTTAACGTTAAAGGT 1138
| | | | |
Db 241 ACCTGGGATGCTAATGCCAATTTAAACCAAGCATGGCTCAACATCTTAACGTTAAAGGT 300
QY 1139 GAGGATGACATGGAGCGCTGTATGTGTTCAGGGTAGAGCTTGGCTAAGCTGTATGGT 1198
| | | | |
Db 301 GAGGATGACATGGAGCGCTGTATGTGTTCAGGGTAGAGCTTGGCTAAGCTGTATGGT 360
QY 1199 GGTGATATTGACACGTTGAAAAAGATTGTGATTTGAGCGCGGTGATGACCGA 1258
| | | | |
Db 361 GGTGATATTGACACGTTGAAAAAGATTGTGATTTGAGCGCGGTGATGACCGA 420
QY 1259 GGTGAAATCTTAACTTCTACATCCGGGTGAATTTACATGGGGTGTGTGCGCCCTTGC 1318
| | | | |
Db 421 GGTGAAATCTTAACTTCTACATCCGGGTGAATTTACATGGGGTGTGTGCGCCCTTGC 480
QY 1319 ATGTACAGCCATCATTTTTCATTTGCTGGGGGATACCTGTATCTCAACAGTACGCGT 1378
| | | | |
Db 481 ATGTACAGCCATCATTTTTCATTTGCTGGGGGATACCTGTATCTCAACAGTACGCGT 540
QY 1379 TCATGATGATGCGCTTGGGGTTGATTTCAACATGATGAGGTTATGTTCTTGGG 1438
| | | | |
Db 541 TCATGATGATGCGCTTGGGGTTGATTTCAACATGATGAGGTTATGTTCTTGGG 600
QY 1439 CTGATGACACAGTATCAACAGGAAAAAGCGGCTTGGCGTATCAACAATCGTCAATGGC 1498
| | | | |
Db 601 CTGATGACACAGTATCAACAGGAAAAAGCGGCTTGGCGTATCAACAATCGTCAATGGC 660
QY 1499 CACATTACCAAGATCACTCGAATTGATGCGGATGTGACGTAACCGTAGCCATTTC 1558
| | | | |
Db 661 CACATTACCAAGATCACTCGAATTGATGCGGATGTGACGTAACCGTAGCCATTTC 720
QY 1559 CCAGCGCCCTCAGTTCATATCATATCCAAAGATTAAACACTGAGAGTTTGGAACTTGG 1618
| | | | |
Db 721 CCAGCGCCCTCAGTTCATATCATATCCAAAGATTAAACACTGAGAGTTTGGAACTTGG 780
QY 1619 GTCACTTTGGATCATTTTGGACGTCAACCGATATCAAGTATCTTCAATACCCG 1678
| | | | |
Db 781 GTCACTTTGGATCATTTTGGACGTCAACCGATATCAAGTATCTTCAATACCCG 840
QY 1679 TTTTCAGTCTAA 1690
| | | | |
Db 841 TTTTCAGTCTAA 852

RESULT 3
US-10-329-960-1
; Sequence 1, Application US/10329960
; Publication No. US2003009277A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag
; FILE REFERENCE: PBI6PI
; CURRENT APPLICATION NUMBER: US/10/329,960
```

```
; CURRENT FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44016)..(44016)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44905)..(44905)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45593)..(45593)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45732)..(45732)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (47036)..(47036)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51334)..(51334)
; OTHER INFORMATION: n equals a, t, g or c
```

```
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (10091)..(10091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119244)..(119244)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
```

```
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152530)..(152530)
```

Query Match: 25.6%; Score 744.2; DB 15; Length 1830121;  
Best Local Similarity 66.7%; Pred. No. 1.4e-232;  
Matches 1134; Conservative 0; Mismatches 538; Indels 27; Gaps 4;

```
QY 2 AGAAGTTGTTATGCTCAGGGTTATCGAGTTCCCAATATTGACCCCGTATTGTT 61
DB 958086 ATAGAGAAATATTGAAATTAATTAATTTACTTCCCACTTGATCCAGATTTT 958145
QY 62 TCGATCGC-----CCTTAGCGGTGCGTGTATGCTTGATGATTTGGTTC 115
DB 958146 ACGCTTGGCGAATGAAATATGAGGCTTACCTTGATAGCTTGATGATCTTTAAGTTT 958205
QY 116 CTTTTCGTATGTGTTGGCAATGCGGAGCGATCGCGGCGGAGTGTGACGCGT 175
DB 958206 GTTTTTCACGTTGCTTGCCTTCGCGTTCGCGTCTAATCCCAATATGCGTTCACAGTA 958265
QY 176 GAGCAAGTCTGACTTGTATTTATGCGCGCTTTTATGAGTATGATCGTGGCGAGTT 235
DB 958266 GATCAAGTTGATAGCTTACTTTTACACGTTTATGAGGAGTGTATTATGCGAGCTGT 958325
QY 236 GGTATGTGATCTTACAAATTTGATCTGTCTTGTGACCTCTTATTATTATCAA 295
DB 958326 GGCATGATTTTCTATATCTGATCATTTCTTACAAAGACCACTTATTATTCGC 958385
QY 296 GTTGTGACTGCGCGCATCTCTTCCACGCGGCTTATGGTGTGATACCGCATGTC 355
```

Db 958386 GTTTGGGAAAGTGAATGTCGTTCAACGGTGGCTTAATGTTGTAATGTTGCTATGAT 958445  
 Qy 356 TGGTATGCGCGTAAACCAACGACCTTCTTGTGTGGCCGATTTTGTGCCCCCTTA 415  
 Db 958446 TGGACATCTTATCTCAAAAAGCTAATTTTGGCAACGGCTGATTTTGTGGCCCTTTG 958505  
 Qy 416 GTGGCATTGCGTTTGGGGAATGGACGTATCGTAACTTTATGAAATGTAACCTTTGGGA 475  
 Db 958506 ATTCGCTTGTGTTAGGTTAGCAGAAATGTAATTTCAATCTTAATCTTGAACATAGGGGA 958565  
 Qy 476 CGAGTAACGGAATGCTTGGGCTTTGTATTCCTTAATGTTGGCC---CACTGGCCGCG 532  
 Db 958566 CCGGAACCAATGTCCTTGGGCAATGATTTTCCCAATGATCCTCTTTTACGCTCGT 958625  
 Qy 533 CATCTTCAAGCTTATGAAATGCGCTTAGAAGGGGTGCTGCTTTATCTTAAT 592  
 Db 958626 CATCAACACACTTATAGAGCTTTTGAAGGCTGTGTGTTTACGATTCGAAT 958685  
 Qy 593 TGGTATTGTTAACTGTCGCTGCTAGGCAAGCTATCCGCACTGTTTATGCTGGATAC 652  
 Db 958686 ATTTTATTAATAAACAAGCTCAATGCGTCTGTGTGCAAGTTATCTTAATGCTTAT 958745  
 Qy 653 GGTACATTCGCTTCTGTGTGAATACGTCCGTGACCAAGTCTCAGTTGGGTCTGTT 712  
 Db 958746 GCGCTCTCCGTTTATTTGGAAATATGTCGTGAACCTGAAG-----TTGAATTTTC 958799  
 Qy 713 GGTGCTTCAATTCANGGGGCAATCCTCTCCTTACCTATGCTATCGTATTTTG 772  
 Db 958800 TTTGGATTTATTAACAGAGGGCAAGCCCTTGTCCCAATGATTAATGCTGTCTTC 958859  
 Qy 773 ATGATGTTTGTCTTCAACAGCCGGTTGTATCAAGCCGTGTACAGCAAAATAGGT 832  
 Db 958860 ATTATGCTTGGCTTATTCAGC-----AAAAGTCGCTCAATAAATAGGAG 958907  
 Qy 833 AGTTAGGTAAACAGTATTTAGATCTTTGTCAAGCATGTCATCAAGGTGTTGGTT 892  
 Db 958908 ATTTTATGAAGCAATATCTGAGCTTTGTCGCGCATTTGTAAGAGGGGAATGGT 958967  
 Qy 893 GAAATGAAGCAACGGGCAAGCTTGTGATCTGTATTAATGATTCAGTACCTACAT 952  
 Db 958968 GCTAATGAAGCAATGAGTAAGCATGCTCAAGCTATTAATGCAATTTAGAAATGAT 959027  
 Qy 953 GTGGGCAACATGATGTTTCTCTAGTACTACACGCAAGGTTTGAAGGCTGCGTA 1012  
 Db 959028 GTGGCAATATATCAATTTCCGCTGATTAACCCGTAATTAATGGAAGGCGGAT 959087  
 Qy 1013 GCGGAGTCTGCGCTAATATCTGTGTTACGATATGCGCGGATTTTGGCAATTAAGT 1072  
 Db 959088 GCTGAATTTTGTAGTATATCTGTGATATGACATGCGCTGATTTTCGCGCACTTGGC 959147  
 Qy 1073 ACCAAACCTGGGATGCTAATGCAATTTAAACAGCATGCTCAACATCCTTACCT 1132  
 Db 959148 ACGAAACCTGGGATGCTAATGCAATTTAAAGAGCTTGGCTTCAATTCGCAATTCGT 959207  
 Qy 1133 AAAGGTGAGATGACATGGAAGCGGTGTATGTTCAAGGTGAGCTTGGGCTAAGCT 1192  
 Db 959208 AAGAGGCTGTATGATATGAGGCGCGTATATGTGTGCAAGGACAGACATGCGTAAGCT 959267  
 Qy 1193 GATGTGTGCTAATTTGACCAAGTTGAAAGATTTGATGATTTGAGCCGTGCGTGTAT 1252  
 Db 959268 AATGAGAAATCTATCATGATCAGTAAGTAATTTGTAATTAACGAAGGATTTGAT 959327  
 Qy 1253 GACCGAGGTGAATTTCTTAATCTTCAATCCGGGTGAATTTCAATGGGGTCTTTGGCC 1312  
 Db 959328 GATAGGAGAGATTTTAACTTTTAAACCTCGGGAAATTTGATCTTGTGTCTTCGT 959387  
 Qy 1313 CCTTGATGTACAGCATATTTTCAATGCTGGGGATACCTTGTATCTCAACAGTACT 1372  
 Db 959388 CCTGTATGATACGATATCTTTTCTGTGTGGGGAATCTTAAATCTTACTACTAT 959447  
 Qy 1373 CAGCGTTCAATGATGTCCTTGGGTTGAATTTCAACAGTGTGAGGTTTATGTTTC 1432  
 Db 959448 CAGCGTTCTGTGATGTTCCGCTGTGATTAATTTCAATCAATTCAGGTGTTTACCTTC 959507

Qy 1433 CTTCGCTGATGACACATCAACAGGAAAGCCGGCTTGGCGTATCAACAGATCGTC 1492  
 Db 959508 TTAGCACTTATGCTCATGATCAACGGGCAAAAAGCCGCAAGCATATCATTAATTTGG 959567  
 Qy 1493 AATGGCACTTTTACCAAGATCACTCGAATTTGATGCGGATGTGCAAGCTTAAACGTGAG 1552  
 Db 959568 AATGGCAATTTTATGAAGATCAGCTTGAGTTAATGCGTACGTAACCTTAAACGCGAG 959627  
 Qy 1553 CCATTCACGCGCCCTCAGTTCATATCAATCCAAAGATTAAACACTGAGATTGGA 1612  
 Db 959628 CTTTCCCAATTACCAAAATTAAGATTAATCCAAATGCAATGCAATTTGGA 959687  
 Qy 1613 ACTTGGGTCACTTGGATTTTGTGACGTCAACGGAATATCAGTTCACAGATCTTATTC 1672  
 Db 959688 ACTTGGGTCAACATGATGATTTTAAAGTGTGCTATCAATCCCAACCAATTA 959747  
 Qy 1673 TACCGTTTCACTTAAAT 1691  
 Db 959748 TATCCTTTTCCGCTCAAT 959766  
  
 RESULT 4  
 US-10-329-670-1  
 ; Sequence 1, Application US/10329670  
 ; Publication No. US20040018503A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Fleischmann et al.  
 ; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag  
 ; FILE REFERENCE: PB186P1  
 ; CURRENT APPLICATION NUMBER: US/10/329,670  
 ; PRIOR FILING DATE: 2002-12-24  
 ; PRIOR APPLICATION NUMBER: US 09/643,990  
 ; PRIOR FILING DATE: 2000-08-23  
 ; PRIOR APPLICATION NUMBER: US 08/487,429  
 ; PRIOR FILING DATE: 1995-06-07  
 ; PRIOR APPLICATION NUMBER: US 08/426,787  
 ; PRIOR FILING DATE: 1995-04-21  
 ; NUMBER OF SEQ ID NOS: 1  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 1  
 ; LENGTH: 1830121  
 ; TYPE: DNA  
 ; ORGANISM: Haemophilus influenzae  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (4747)..(4747)  
 ; OTHER INFORMATION: n equals a, t, g or c  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (921)..(921)  
 ; OTHER INFORMATION: n equals a, t, g or c  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (10150)..(10150)  
 ; OTHER INFORMATION: n equals a, t, g or c  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (29298)..(29298)  
 ; OTHER INFORMATION: n equals a, t, g or c  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (36543)..(36543)  
 ; OTHER INFORMATION: n equals a, t, g or c  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (36551)..(36551)  
 ; OTHER INFORMATION: n equals a, t, g or c  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (36636)..(36636)  
 ; OTHER INFORMATION: n equals a, t, g or c

FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (40808)..(40810)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (44416)..(44416)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (44905)..(44905)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (44975)..(44975)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (45593)..(45593)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (45732)..(45732)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (47036)..(47036)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (51334)..(51334)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (51602)..(51602)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (51786)..(51786)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (51805)..(51805)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (55369)..(55369)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (65309)..(65309)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (65313)..(65313)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (80024)..(80024)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (100691)..(100691)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (102696)..(102696)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (105121)..(105121)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:

NAME/KEY: misc\_feature  
LOCATION: (107248)..(107248)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (117136)..(117136)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (119750)..(119750)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (119924)..(119924)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (120038)..(120038)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (121344)..(121344)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (122167)..(122167)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (122336)..(122336)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (131340)..(131340)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (131360)..(131360)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (139910)..(139910)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (140398)..(140398)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (142750)..(142750)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (145058)..(145058)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (145171)..(145171)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (145942)..(145942)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (150841)..(150841)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature



;; CURRENT APPLICATION NUMBER: US/10/282,122A  
;; CURRENT FILING DATE: 2003-02-20  
;; PRIOR APPLICATION NUMBER: 60/191,078  
;; PRIOR FILING DATE: 2000-03-21  
;; PRIOR APPLICATION NUMBER: 60/206,848  
;; PRIOR FILING DATE: 2000-05-23  
;; PRIOR APPLICATION NUMBER: 60/207,727  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: 60/230,335  
;; PRIOR FILING DATE: 2000-09-06  
;; PRIOR APPLICATION NUMBER: 60/230,347  
;; PRIOR FILING DATE: 2000-09-09  
;; PRIOR APPLICATION NUMBER: 60/242,578  
;; PRIOR FILING DATE: 2000-10-23  
;; PRIOR APPLICATION NUMBER: 60/253,625  
;; PRIOR FILING DATE: 2000-11-27  
;; PRIOR APPLICATION NUMBER: 60/257,931  
;; PRIOR FILING DATE: 2000-12-22  
;; PRIOR APPLICATION NUMBER: 60/267,636  
;; PRIOR FILING DATE: 2001-02-09  
;; PRIOR APPLICATION NUMBER: 60/269,308  
;; PRIOR FILING DATE: 2001-02-16  
;; Remaining Prior Application data removed - See File Wrapper or PALM.  
;; NUMBER OF SEQ ID NOS: 78614  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 30692  
;; TYPE: DNA  
;; LENGTH: 852  
;; ORGANISM: Pasteurella multocida  
US-10-282-122A-30692

Query Match 17.4%; Score 505.4; DB 13; Length 852;  
Best Local Similarity 74.6%; Pred. No. 7.8e-156;  
Matches 635; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

QY 840 TGAACGATTTAATCTTTGTCAAGCGATGTCATCAAGTCTTTGGTTGAATG 899  
DB 2 TGAACGATTTAATCACTTGGCAACGATCTGTGATGAAGTCTTTGGTTGAATG 61  
QY 900 AACGACGGGCAAGGCTTTGTTGACTGTGATTAAATCCGATTGACCTTACATGAGGCA 959  
DB 62 CACGACCTGGCAAAAGCTGTCTGACCGTATGATGCGGATCTACCTTACATGTTGCCA 121  
QY 960 ACAATCACTTCTCTAGTACTACACGCAAGATTTTGGAAAGCTGCCGAGT 1019  
DB 122 ATTAACCAATTCCTGTTGATCACCAACGCAAAAGCTATTTGAAAGCGCAATTCAGAT 181  
QY 1020 TCGTGGCTATTTGTTGTTACGATATGCGCGGATTTTCCCAATTAAGTACCAAA 1079  
DB 182 TTTTAGGTTAATTCGTGGCTATGCAATGCGGCAATTTCCGCAACCTCGCACCAAAA 241  
QY 1080 CTTGGAGTCTAATGCAATTTAAACCAAGATGGCTCAACAATCTTACCGTAAAGTG 1139  
DB 242 CTTGGAGTCCATCTAATGAAACAGTGCATGCTCAATATCCGACCGCAAGGCA 301  
QY 1140 AGGATGACATGGGACCGCTGATGCTTCAAGGCTAGAGCTTGGCTTAAAGCTGATG 1199  
DB 302 CCGATGATATGGGAGCTGTTATGCGTAAAGGCGGATGCGCAACCGAAGCGTG 361  
QY 1200 GTCAATTTGACGATGAAAAAGATTTGATGATTTGAACCGTGGCTGATGACCGAG 1259  
DB 362 AAACCGTATCAATTTGGCAAAATCGTCAATATCTGCGCAACGCAATTTGATGAC 421  
QY 1260 GGAATCTTAATCTTACATCGGGGATTTCAATGCGGCTTGGCGCTTGA 1319  
DB 422 GGAATATGACCTTCTTCAACCCCGGGAATTTGAATGAGTGGTGGCTTCACTCTTCA 481  
QY 1320 TGTACGCAATCTTTTCAATGCTGAGGAGATCTTGTATCTCAACAGTACAGCTT 1379  
DB 482 TGCACACCCACCTTCTCATCTCGGATGATACCTCTATCTACAGATTAAGCGTT 541  
QY 1380 CATGATGATGCGCTTGGGTTGAATTTCAACATGAGTGGATTTATGTTCTTGGCG 1439

DB 542 CTTGATGATGCCCGCTTGGCTTGAATTTCAACCAATTTCAAGTTCACCTTTCTGCT 601  
QY 1440 TGATGGCAAGATCAAGGGAAGAAAGCGGCTTGGCTATCAAGATCTCATAGGCG 1499  
DB 602 TGATGGCAAAATTAACGGGCAAAAGCGGTTAAGCGGTACCAAAATCATCATAGCAC 661  
QY 1500 ACATTTACCAAGATCACTCGAATTTGATGCCGATGTGACACTTAAAGTACCATTC 1559  
DB 662 ACATTTAAGAAATCACTTATCTCATGAAATATGCAATTTAAAGTACCCCTTCC 721  
QY 1560 CAGGCGCTCAGTTCATCATCATCAATGCAAAAGATTTAAACCTGACAGATTTGAAA 1619  
DB 722 CTTGCCCAATTTACATTTAACCCGAGATTTAAACCTGAAAGATTTAGAACTTGG 781  
QY 1620 TCACCTTGATGATTTTGAAGCTCACCGGATATCAATGCTTCAAGATCTTATCAAT 1679  
DB 782 TGACATGATGATTTTAAAGTTCACCGGCTATCATGACGATGCAATTAATATCTT 841  
QY 1680 TTTCACTTAA 1690  
DB 842 TCTCTGTGTAA 852

## RESULT 6

US-10-282-122A-22162

; Sequence 22162, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlson, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trewick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: EUTRA.034A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 22162

; LENGTH: 852

; TYPE: DNA

; ORGANISM: Haemophilus influenzae

US-10-282-122A-22162

Query Match 15.9%; Score 462.2; DB 13; Length 852;  
 Best Local Similarity 71.4%; Pred. No. 1.7e-141;  
 Matches 608; Conservative 0; Mismatches 243; Indels 0; Gaps 0;

QY 840 TGAACAGATTTAGATCTTTGTCAAGCGATGTCATCAAGTGTGGTTGAAATG 899  
 DB 2 TGAACCAATCTTGGCTTTGTCCGCCATTTGTAGTAAGGGGAATGGTTGCTAAAG 61

QY 900 AACGAACGGGCAAGGCTTTGTGACGTGATTAATGCGATTTGACCTACAGATGGGCA 959  
 DB 62 AACGTAACGATTAAGCACTTCTCAAGCTCAATTAATGCAATTAAGTAATGATGGCA 121

QY 960 ACAATCAGTTCTCTAGTACTACAGCAAGGTTTGGAAAGTGGCCGAGCGAGT 1019  
 DB 122 ATAAATCAATTTCCGGGATTAATACCCGTAAGTAATGAAAGGGGAGTTGCTGAT 181

QY 1020 TECTGGCTAATTTGCTGTATACGATATGCGGCGATTTTGGCCAAATGATCAAAA 1079  
 DB 182 TTTTGGTTATTTCTGTGATATGACAAATGCGCTGATTTCCGCCACTTGGCAGCAAAA 241

QY 1080 CCTGGAGTCTAATGCAATTTTAAACCAAGCATGCTCAACATCTTACCTTAAAGTG 1139  
 DB 242 CTGGAGATCTAATGCAATTTTAAAGCAAGCTTGGCTTCAATTCGCAATCGTAGAGCG 301

QY 1140 AGGATGACATGGGACGCGGTATGCTTCAAGGTAGAGCTTGGCTAAGCTGATG 1199  
 DB 302 TTGATGATATGGGGCCGATATATGCTGTCAGAGGACAGACATGGCTAATGAG 361

QY 1200 GTCATATTTGACAGTTGAAAAAGATTTGATGATTTGAGCCGTGGCTGATGACGAG 1259  
 DB 362 AAACATGATGATGACGTAACGTAATTTGTAATTAACGTAATGATGATGATGAG 421

QY 1260 GTGAATTTCTTAATCTTCAATCCGGGTGATTTTCAACATGGGGTTCGCCCTTGA 1319  
 DB 422 GAGAAATTTTAACTTTTAACTTGGGGAATTTGATCTTGGTGTCTTCTTGTGA 481

QY 1320 TGTACAGCCATTTTCTTATGCTGGGGATACCTTGTATCAACAGTACTACAGCTT 1379  
 DB 482 TGCATACGATCTTTTCTTCTTGTGGGATACCTTATCACTTACTAGCTACAGCTT 541

QY 1380 CATGTATGTCCTTGGGGTGAATTTTCAACATGGGTGAGTTATGTTCTTGGC 1439  
 DB 542 CCGTATGATGTCGGCTTGGATTTGAATTTCAATCAATTCAGGTTTACTTCTTGA 601

QY 1440 TGATGCGACATTCACAGGAAAAAGCCGGCTTGGCGATTCACAAAGTCTCATGCGC 1499  
 DB 602 TTAATGCTCAGATCAAGGCAAAAAAGCCGCAAGGCAATATCAATTAATTTGAAATGCGC 661

QY 1500 AATTTACCAAGATCAATGCAATGATGATGCGGATGAGGCTAAAGCGTAAAGCCATTC 1559  
 DB 662 ATATTTATGAATCACTTGAATTAATGCTGATCACTTAAACCGGAGCTTTTC 721

QY 1560 CAGCCCTCAGTTCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 1619  
 DB 722 CATTCACAAATTTAGAAATTAATCAAGATTAATAAAGCTTGAAGTTTGGAACTTGG 781

QY 1620 TCATCTTGGATTTTGAAGTCACCGGATTCAGTTCAGATCTTCAATACCGCT 1679  
 DB 782 TCACATGATGATTTTAAAGTCGTTGCTATCAATCCAGCAACCAATTAATATCTT 841

QY 1680 TTTCACTTAA 1690  
 DB 842 TTTCCGCTTAA 852

RESULT 7  
 US-10-282-122A-32325  
 ; Sequence 32325, Application US/10282122A  
 ; Publication No. US20040029129A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, Liangau  
 ; APPLICANT: Zamudio, Carlos  
 ; APPLICANT: Malone, Cheryl

APPLICANT: Haselbeck, Robert  
 APPLICANT: Ohlsen, Kari  
 APPLICANT: Zyskind, Judith  
 APPLICANT: Wall, Daniel  
 APPLICANT: Trawick, John  
 APPLICANT: Carr, Grant  
 APPLICANT: Yamamoto, Robert  
 APPLICANT: Forayth, R.  
 APPLICANT: Xu, H.  
 TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
 FILE REFERENCE: ELITRA.034A  
 CURRENT APPLICATION NUMBER: US/10/282,122A  
 PRIORITY FILING DATE: 2003-02-20  
 PRIOR APPLICATION NUMBER: 60/191,078  
 PRIOR FILING DATE: 2000-03-21  
 PRIOR APPLICATION NUMBER: 60/206,848  
 PRIOR FILING DATE: 2000-05-23  
 PRIOR APPLICATION NUMBER: 60/207,727  
 PRIOR FILING DATE: 2000-05-26  
 PRIOR APPLICATION NUMBER: 60/230,335  
 PRIOR FILING DATE: 2000-09-06  
 PRIOR APPLICATION NUMBER: 60/230,347  
 PRIOR FILING DATE: 2000-09-09  
 PRIOR APPLICATION NUMBER: 60/242,578  
 PRIOR FILING DATE: 2000-10-23  
 PRIOR APPLICATION NUMBER: 60/253,625  
 PRIOR FILING DATE: 2000-11-27  
 PRIOR APPLICATION NUMBER: 60/257,931  
 PRIOR FILING DATE: 2000-12-22  
 PRIOR APPLICATION NUMBER: 60/267,636  
 PRIOR FILING DATE: 2001-02-09  
 PRIOR APPLICATION NUMBER: 60/269,308  
 PRIOR FILING DATE: 2001-02-16  
 Remaining Prior Application data removed - See file wrapper or PALM.  
 NUMBER OF SEQ ID NOS: 78614  
 SOFTWARE: Patent in version 3.1  
 SEQ ID NO 32325  
 LENGTH: 849  
 TYPE: DNA  
 ORGANISM: Proteus mirabilis  
 US-10-282-122A-32325

Query Match 13.9%; Score 403.8; DB 13; Length 849;  
 Best Local Similarity 67.3%; Pred. No. 4e-122;  
 Matches 570; Conservative 0; Mismatches 277; Indels 0; Gaps 0;

QY 840 TGAACAGATTTAGATCTTTGTCAAGCGATGTCATCAAGTGTGGTTGAAATG 899  
 DB 2 TGAACCAATCTTGGCTTTGTCCGCCATTTGTAGTAAGGGGAATGGTTGCTAAAG 61

QY 900 AACGAACGGGCAAGGCTTTGTGACGTGATTAATGCGATTTGACCTACAGATGGGCA 959  
 DB 62 AACGTAACGAAACCCGTTGTTTAAAGGATTAATGCGATTCGAAATTAATGTTGCCA 121

QY 960 ACAATCAGTTCTCTAGTACTACAGCAAGGTTTGGAAAGTGGCCGAGCGAGT 1019  
 DB 122 ATAAATCAATTTCCAGTACTACAGCAAGCTTAAAGTTTTCAAAGCCGAGTTGGCAAC 181

QY 1020 TECTGGCTAATTTGCTGTATACGATATGCGGCGATTTTGGCCAAATGATCAAAA 1079  
 DB 182 TTTTGGTTATTTCTGTGATATGACAAATGCGCTGATTTCCGCCACTTGGCAGCAAAA 241

QY 1080 CTTGGAGTCTAATGCAATTTTAAACCAAGCATGCTCAACATCTTACCTTAAAGTG 1139  
 DB 242 CTGGAGATCTAATGCAATTTTAAAGCAAGCTTGGCTTCAATTCGCAATCGTAGAGCG 301

QY 1140 AGGATGACATGGGACGCGGTATGCTTCAAGGTAGAGCTTGGCTAAGCTGATG 1199  
 DB 302 TTGATGATATGGGGCCGATATATGCTGTCAGAGGACAGACATGGCTAATGAG 361

QY 1200 GTCATATTTGACAGTTGAAAAAGATTTGATGATTTGAGCCGTGGCTGATGACGAG 1259  
 DB 362 GCGACTTGTATTAATGATGATTAAGTGTGATTAATCTTAATCAACGATTAATGATGAG 421

QY 1260 GTGAATTCCTTAACCTTCTACATCCGGGTGAATTTACATGGGGTGTGGCCCTTCA 1319  
 DB 442 GTGAGATGTGACCTTTTATATATCCGGAGAAACAGCGTTAGGGTGTAGCTCATCA 481  
 QY 1320 TGTACAGCCATCTTTTTCATTTGCTGGGGGATACCTTGTATCTCAACAGTACTGCGCTT 1379  
 DB 482 TGCATACACATACCTTTTTCATTTGCTGGGTGACAGGCTTTATTTAACCTCATATACAGTA 541  
 QY 1380 CATGATGATGTCCTTGGGGTGAATTTCAACATGTCAGAGTTATGTTCCCTGGGC 1439  
 DB 542 GCTGTGATGTCCTTATAGGGTGTGAATTTTATCAATTCATGCTTTGATTTATAGCGC 601  
 QY 1440 TGATGCAACATCAACAGGAAACCCGGCTTGGCGATCAACAGATCGCAATGCGC 1499  
 DB 602 TGTGTGCTCAATCAACAGGAGATTAACAGGTAAAGCATTCATTAAGATTAATGCTC 661  
 QY 1500 ACATTTACCAAGATCACTCCGATTTGATGCGGATGTCAGCTAAACGTGAGCCATTCC 1559  
 DB 662 ACATTTATGAGAACCAATTAACCTGATGCGGATGTTCACTAAAGAGCCATTGCG 721  
 QY 1560 CAGGCGCTCAGTTTCATTCATCCAAAGATTAAACACTGACGATTTGGAACTTGGG 1619  
 DB 722 CATTAACCTAAGCTACATATTTATCCAAACATTTAAGACATTTAGATATATAGAACTTGGG 781  
 QY 1620 TCACCTTGGATTTTGGACGTCACCGGATATCAGTTCCAGATTCCTTAATCAATCCCGT 1679  
 DB 782 TCACGACAGATGACTTTAGTGTGAAGTTATCACTGTCAAGAGCGATTAATAATCCCGT 841  
 QY 1680 TTTTCACT 1686  
 DB 842 TCACCGT 848

## RESULT 8

US-09-741-669-182  
 ; Sequence 182, Application US/09741669  
 ; Patent No. US2002022718A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Forsyth, R. Allyn  
 ; APPLICANT: Ohlsen, Kari L.  
 ; APPLICANT: Zyskind, Judith W.  
 ; TITLE OF INVENTION: Genes identified as required for  
 ; FILE REFERENCE: ELITRA.009A  
 ; CURRENT APPLICATION NUMBER: US/09/741,669  
 ; CURRENT FILING DATE: 2000-12-19  
 ; PRIOR APPLICATION NUMBER: US 60/173005  
 ; PRIOR FILING DATE: 1999-12-23  
 ; NUMBER OF SEQ ID NOS: 481  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 182  
 ; LENGTH: 876  
 ; TYPE: DNA  
 ; ORGANISM: Escherichia coli  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)...(876)  
 US-09-741-669-182

Query Match 8.5%; Score 248.2; DB 9; Length 876;  
 Best Local Similarity 58.8%; Pred. No. 1.7e-70;  
 Matches 510; Conservative 0; Mismatches 288; Indels 69; Gaps 2;  
 QY 14 ATGCTCAGGGTATCTGCAAGTTCCCAATATGACCCCGATTTGTTTCATCGGCCCT 73  
 DB 1 ATGACACAGTATCTCATTTCCGAGATTGATCCGGTATTTCTCAATAGAACCC 60  
 QY 74 CTAGCGGTGCGGTGATGAGCTGATATTGTTGGGTTTCTTTTGGATAGGTTG 133  
 DB 61 GTGGGCTTCACTGTATAGCGGCTGATGATCTGAGGTTTTCATTTTGGCAATGTTGGCTG 120  
 QY 134 GCCAATCCCGAGCGGATGCGGGCAGTGGTTGAGCGGTGACAGTCTGTACTTG 193

DB 121 GCAACACACCGGGGAATCGCCGGGACGCGCTGACCAAAAAATGAAGTGAATAACTTA 180  
 QY 194 TTAATCGCGGCTTTTATAGGTATGATCGGTGGCCAGTTGGTTATGATCTTAC 253  
 DB 181 CTATAGGGGCTTCTCGGCGTCTTCTCGGGGACATTAATGTTATGTTCTTAC 240  
 QY 254 AATTTATCTGTTCTTGTGACCCCTTTATTTATTTCAAGTGTGAGCGGCGATG 313  
 DB 241 AATTTCCGAGTTTATAGCGCATCCGCTGATCTGTTCCGTTCTGGACGCGGCGATG 300  
 QY 314 TCTTTCCAGCGGCTTATTTGGGTGTATACCGCATGTTCTGGTATGCGGCTTAAAC 373  
 DB 301 TCTTTCCAGCGGCTGATTTGGGTTATCGTGTATGATTAATCTTCGCGCGGCTACT 360  
 QY 374 CAAGCACCTTTTGTGGGCGGATTTGTTGCCCCCTTATAGTCCATGCGTTGGGG 433  
 DB 361 AAGGTTCTTCTTTCAGAGTCTGATTTTATCGACACATCATTTCCGTTGGTGTGT 420  
 QY 434 ATGGACGTATCGGTAACTTTATGAATAGTAACTTTGGGACGAGT----- 480  
 DB 421 GCCGGGCTGTGGCACTTTATTAAGGTAAATTTGGGGCCGCGTTGACCCGAATTC 480  
 QY 481 -----AAGGAT 487  
 DB 481 CCGTTGCAATCTGTCCCTGCTCCGTAACAGAAATATTTGCTGCAAAACCAAC 540  
 QY 488 GTGCGTGGGCTTTTATTCCTTAATGTTGCCACATGCGGCGCAATCTTCAAGCTT 547  
 DB 541 CCGAGTGGCAATCATTTTGACACCTTACGCTGTCTGCGCGCACCACTCAAGCTT 600  
 QY 548 TATGAATTCCTTAGAAGCGGTGCTGTTCTTTATTTCTTAATTTGTTATTTGTA 607  
 DB 601 TACGAGTGTGCTGGAAGGTGTGTGTGTTATTTATCTCAACCTGATATTCGTAA 660  
 QY 608 CCTGTCCGCTAGGACGCGTATCCGAGCTTTTATGCTGATACGTTACTTCCGCTTC 667  
 DB 661 CCAGGCCCAATGGGAGCTGTCTCAGGTTGTTCTGATTTGTTGCGGCGCTTTCGCAATC 720  
 QY 668 CTGTGGAATACGTCGCGGACCGAGATGCTAGTTG---GGTCTGTTGGTGGCTCATTT 724  
 DB 721 ATGTGATGTTTTCGACAGCCGACGCGAGTTTACCGGTGCTGGGTGACGATATC 780  
 QY 725 TCAATGGGCAAACTCTCTTACTATGATGATCAATCGGTATTTGATGATGTTGG 784  
 DB 781 AGCATGGGCAATTTCTTTCATCCGATGATTTGTCGGGTGATCATGATGATGTCGG 840  
 QY 785 TCTTACAGCGCGGTTGTATCAAGAC 811  
 DB 841 GCATATGTCGACACCCACACCAACAC 867

## RESULT 9

US-09-790-988-1  
 ; Sequence 1, Application US/09790988  
 ; Patent No. US20020127687A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: SHIGENOBU, SHUJI  
 ; APPLICANT: MATSUMAE, HIDEKI  
 ; APPLICANT: HATTORI, YOSHIAKIRA  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS  
 ; FILE REFERENCE: 081356/0159  
 ; CURRENT APPLICATION NUMBER: US/09/790,988  
 ; CURRENT FILING DATE: 2001-02-23  
 ; PRIOR APPLICATION NUMBER: JP2000-107160  
 ; NUMBER OF SEQ ID NOS: 7  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 1  
 ; LENGTH: 640681  
 ; TYPE: DNA  
 ; ORGANISM: Buchnera sp.



US-10-297-465A-1

Query Match 4.0%; Score 115.4; DB 17; Length 2731748;  
Best Local Similarity 53.8%; Pred. No. 8e-24;  
Matches 261; Conservative 0; Mismatches 221; Indels 3; Gaps 1;

```
OY 36 TTCCCAATATGACCCCGATATGTTTTCGATCGCCCTCTAGCCGTGCGGTGATGGCT 95
DB 2214488 TCACGCTATCGATCCGATGCTTCTCACTTGTCGGGTGAAGGTGATGATGCT 2214429
OY 96 TGAATGATTTGGGTGCTTCTTTTTCATGTTGTTGCTTCCCAATGCGCGGATCCCG 155
DB 2214428 TGATGATTTTGGCAATTTTGGGCGCGGTGCTTGGAGCCCAACGCAATCCAGCCG 2214369
OY 156 CGGCGAG--TGGTGAACGCGGTGAGCAAGTCTGACTGTTTATTCGCGGCTTTTAA 212
DB 2214368 GGCCTTGTGCTGTGTCAACATGATGTTTCCGACTGTGTTCTTAACGGAATGATG 2214309
OY 213 GTTAGTATCGGTGCGCGAGTTGTTATGATCTTCACAATTTTGAATCTGTTCTTG 272
DB 2214308 GTGTGTTGTTGGGTGCGGTGCGGTGATGATGCTGTTCAATGCTTCAATGACTTCTG 2214249
OY 273 CTGACCTCTTATTTATTTCAAAAGTGTGACTGCGGCAATGCTTCCACGCGGCTTAT 332
DB 2214248 AAGAACCAATGCTGCTGTTTCAAGTGTGGAAGGCGGCAATGAGCTTTTCAATGAGGTTTGA 2214189
OY 333 TGGGTGATATCACCGCATGTTCTGTTATGCGGTAATAAACCAACGCACTTCTTGGTG 392
DB 2214188 TTGGGTCTCTATGCAAGTGTGCTTGTGAGTGTGCGGTGATGCTTCAAGATGTTGAT 2214129
OY 393 TGGCGATTTTGTGTCCTTTTATGTCATTCGTTTGGGATGGAACGTATCGTAACT 452
DB 2214128 TGCTGATTTTGGCGCACCACTGTCCTCGTGGGCTTGGGTTTGTGGCTTGGCACT 2214069
OY 453 TTATGAATATGAACTTTTGGGAGCAAGTATGCTTGGGCTTTTGAATTCCTTA 512
DB 2214068 TCATGCGCGGTGAGCTTTGGGGAAGTCAACCAATGCTGGGAGTAATTTTCCGC 2214009
OY 513 ATGAT 517
DB 2214008 GTGCT 2214004
```

RESULT 12  
US-10-152-319A-1589/c

/ Sequence 1589, Application US/10152319A  
/ Publication No. US20040072160A1

/ GENERAL INFORMATION:

/ APPLICANT: Mendrick, Donna

/ APPLICANT: Porter, Mark

/ APPLICANT: Johnson, Kory

/ APPLICANT: Higgs, Brandon

/ APPLICANT: Casle, Arthur

/ TITLE OF INVENTION: Molecular Toxicology Modeling

/ FILE REFERENCE: 44921-5089-US

/ CURRENT APPLICATION NUMBER: US/10/152,319A

/ CURRENT FILING DATE: 2002-05-22

/ PRIOR APPLICATION NUMBER: US 60/292,335

/ PRIOR FILING DATE: 2001-05-22

/ PRIOR APPLICATION NUMBER: US 60/297,523

/ PRIOR FILING DATE: 2001-06-13

/ PRIOR APPLICATION NUMBER: US 60/298,925

/ PRIOR FILING DATE: 2001-06-19

/ PRIOR APPLICATION NUMBER: US 60/303,810

/ PRIOR FILING DATE: 2001-07-10

/ PRIOR APPLICATION NUMBER: US 60/303,807

/ PRIOR FILING DATE: 2001-07-10

/ PRIOR APPLICATION NUMBER: US 60/303,808

/ PRIOR FILING DATE: 2001-07-10

/ PRIOR APPLICATION NUMBER: US 60/315,047

/ PRIOR FILING DATE: 2001-08-28

/ PRIOR APPLICATION NUMBER: US 60/324,928

/ PRIOR FILING DATE: 2001-09-27  
/ PRIOR APPLICATION NUMBER: US 60/330,867  
/ PRIOR FILING DATE: 2001-11-01  
/ PRIOR APPLICATION NUMBER: US 60/330,462  
/ PRIOR FILING DATE: 2001-10-22  
/ Remaining Prior Application data removed - See File Wrapper or PALM.  
/ NUMBER OF SEQ ID NOS: 2221  
/ SOFTWARE: Patent Ver. 2.1  
/ SEQ ID NO 1589

/ LENGTH: 2440  
/ TYPE: DNA  
/ ORGANISM: Rattus norvegicus

/ FEATURE:  
/ OTHER INFORMATION: Genbank Accession No. NM\_013030

US-10-152-319A-1589

Query Match 2.9%; Score 84; DB 12; Length 2440;  
Best Local Similarity 58.5%; Pred. No. 1.1e-15;  
Matches 166; Conservative 0; Mismatches 115; Indels 3; Gaps 1;

```
OY 2480 TCTAGAGGTAGAAAGATGACGACCGCAGAAATTGAAAAATCGTGAATGTCACATG 2539
DB 766 TCAGGGGCAAGACAGACCAAGACAGACGCAATTAAGCATATGACATGTCCTCCCT 707
OY 2540 CGGAAAGCAGCGGAACTCTTTTACAGCGCATATGCGCAAGGCTGACGAGATTTG 2599
DB 706 GCAAAAGCCCGCTGAATGATGCTCTGTCCTCCCGCTGCATCAGGGCCCAATGATGTTG 647
OY 2600 GTCAAGATATGAAATTTTGGACCCCATCAATAGAAATCGGG--TTTCAACCGGT 2656
DB 646 GTGAGAGGTTCCTGATTTGAGCCCAATGATTTGAAATGCGGAGCTCACTCCAC 587
OY 2657 AACCAAGGAGACGAGACCAATATATAGATGACCGTCTGAGATTTGATGAT 2716
DB 586 AAGCAGAGAGACCATCTGACATATGATGATGATGATGATGATGATGATGATGATGAT 527
OY 2717 GCCGTGCACTAAACATCATCATCATCTTCAATTTGGTGGGA 2760
DB 526 ACTGTCACACGAGATTCACACACGAGCCCTGCATCGGTTTGA 483
```

RESULT 13

US-10-152-319A-1988/c

/ Sequence 1988, Application US/10152319A  
/ Publication No. US20040072160A1

/ GENERAL INFORMATION:

/ APPLICANT: Mendrick, Donna

/ APPLICANT: Porter, Mark

/ APPLICANT: Johnson, Kory

/ APPLICANT: Higgs, Brandon

/ APPLICANT: Casle, Arthur

/ APPLICANT: Elashoff, Michael

/ TITLE OF INVENTION: Molecular Toxicology Modeling

/ FILE REFERENCE: 44921-5089-US

/ CURRENT APPLICATION NUMBER: US/10/152,319A

/ CURRENT FILING DATE: 2002-05-22

/ PRIOR APPLICATION NUMBER: US 60/292,335

/ PRIOR FILING DATE: 2001-05-22

/ PRIOR APPLICATION NUMBER: US 60/297,523

/ PRIOR FILING DATE: 2001-06-13

/ PRIOR APPLICATION NUMBER: US 60/298,925

/ PRIOR FILING DATE: 2001-06-19

/ PRIOR APPLICATION NUMBER: US 60/303,810

/ PRIOR FILING DATE: 2001-07-10

/ PRIOR APPLICATION NUMBER: US 60/303,807

/ PRIOR FILING DATE: 2001-07-10

/ PRIOR APPLICATION NUMBER: US 60/303,808

/ PRIOR FILING DATE: 2001-07-10

/ PRIOR APPLICATION NUMBER: US 60/315,047

/ PRIOR FILING DATE: 2001-08-28

/ PRIOR APPLICATION NUMBER: US 60/324,928

/ PRIOR FILING DATE: 2001-09-27

/ PRIOR APPLICATION NUMBER: US 60/330,867

```

; PRIOR FILING DATE: 2001-11-01
; PRIOR APPLICATION NUMBER: US 60/330,462
; PRIOR FILING DATE: 2001-10-22
; Remaining Prior Application data removed - See File Wrapper or PAM.
; NUMBER OF SEQ ID NOS: 2221
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1988
; LENGTH: 3950
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. NM_053380
US-10-152-319A-1988

```

```

Query Match      2.7%; Score 77.8; DB 12; Length 3950;
Best Local Similarity 56.7%; Pred. No. 1.9e-13;
Matches 164; Conservative 0; Mismatches 122; Indels 3; Gaps 1;

```

```

QY 2480 TCTAGAGTAGAAGATCAGACCCGAGAGATTGAAAAATCGTGATGTGACACTG 2539
    |||||
DB 724 TCCAGAGTAGAAGACCAACAGAGAGCCAGTTGAAGAATGACAGAGAGTCTCT 665
    |||||
QY 2540 GCGAAGACGCGGAACTCTTCTTTACAGCCGCAATGSCCAAGCTGACAGAGATTG 2599
    |||||
DB 664 GCAAAATGCCCTTCTGAACTCATTTCTCTCTGCTGCATGAATGCCAATTCGTG 605
    |||||
QY 2600 GTACAGTAGTACCATATTTGGACCCATCATCATAGTAAGATGCGGTTTCAAC--CGGT 2656
    |||||
DB 604 GTAGTAGAGTCCCGATGTTAGCAACCATGATGATGGAGAGCGGCGACTGACAGC 545
    |||||
QY 2657 AACCAACGCGAAGACAGACCAACATATATAGAGTACCGCTTGAGATTGAATCACT 2716
    |||||
DB 544 AAGAGAGAGGCAACATGCTGAGAGATGAGAGAGAGTCTGAGCTTGGACCAATG 485
    |||||
QY 2717 GCCGTTCCTTAACCAATCATCTCTGCAATTTGGGTGGAGCAA 2765
    |||||
DB 484 ACTGTACACGACACCCCTATCAAGTCCAGCCACGAGATTGACATTA 436
    |||||

```

```

RESULT 14
US-10-276-774-1331/c
; Sequence 1331, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang, Y. Tom et al
; TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-030
; CURRENT APPLICATION NUMBER: US/10/276,774
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 1331
; LENGTH: 2208
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-276-774-1331

```

```

Query Match      2.6%; Score 75.4; DB 13; Length 2208;
Best Local Similarity 56.0%; Pred. No. 7.6e-13;
Matches 164; Conservative 0; Mismatches 126; Indels 3; Gaps 1;

```

```

QY 2476 CATTTCTAGAGTAGAAGATCAGACCCGAGAGATTGAAAAATCGTGATGTGAC 2535
    |||||
DB 738 CACCTTCACGCGGCAAGACCAACAGAGCCAGTTGAAGAATCATGACAGTGGC 679
    |||||
QY 2536 ACTGCGAAGACGCGGAACTCTTCTTTACAGCGCATATGAGCCAGGCTGACGAGT 2595
    |||||
DB 678 TCTGCAAAAGCTTTCTGAACTCTCGATCTCTCCACCTGCAATGAGCGCAACATAGT 619
    |||||

```

```

QY 2596 ATTGGTCAAGTAGTACCAATATTGGACCCATCACCATAGGAATGCGGTTTCAACG- 2654
    |||||
DB 618 GTTGGTAGTGAAGTTCATTTGGCCCCCATGATTAATGGAGTAGGAGCCCAACAGT 559
    |||||
QY 2655 -GTAACCAACCGGCAAGACCAACATTAATTAAGTAGTACCGCTTGAGATTGAAT 2712
    |||||
DB 558 GAGCAATGAAGAGGACACCATGCTGCAACAGTAGACGTTGAGTGTGAGACTCTGCAC 499
    |||||
QY 2713 CAGTCCGTTGGCCTTAACCATCATCATCTCTGCAATTTGGGTGGAGCAA 2765
    |||||
DB 498 CAAGACGCTACACACACCCGATCATCAGCCCAACAAAGGTTGACATTA 446
    |||||

```

```

RESULT 15
US-10-173-999-149/c
; Sequence 149, Application US/10173999
; Publication No. US20040005563A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David H.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Ovarian Cancer, Compositions
; TITLE OF INVENTION: and Methods of Screening for Modulators of Ovarian
; FILE REFERENCE: 018501-002420US
; CURRENT APPLICATION NUMBER: US/10/173,999
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: US 60/299,234
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/315,287
; PRIOR FILING DATE: 2001-08-27
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/372,246
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 163
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 149
; LENGTH: 2280
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-173-999-149

```

```

Query Match      2.6%; Score 75.4; DB 16; Length 2280;
Best Local Similarity 56.0%; Pred. No. 7.7e-13;
Matches 164; Conservative 0; Mismatches 126; Indels 3; Gaps 1;

```

```

QY 2476 CATTTCTAGAGTAGAAGATCAGACCCGAGAGATTGAAAAATCGTGATGTGAC 2535
    |||||
DB 771 CACCTTCACGCGGCAAGACCAACAGAGCCAGTTGAAGAATCATGACAGTGGC 712
    |||||
QY 2536 ACTGCGAAGACGCGGAACTCTTCTTTACAGGCAATATGSCCAAGGCTGACGAGT 2595
    |||||
DB 711 TCTGCAAAAGCTTTCTGAACTCTCGATCTCTCCACCTGCAATGAGGCAACATAGT 652
    |||||
QY 2596 ATTGTACAGTAGTACCAATATTGGCAACCATCATCATAGGAATCGCGCTTCAACG- 2654
    |||||
DB 651 GTTGGTAGTGAAGTTCATTTGAGTGGCCCCCATGATTAATGGAGTAGGAGCCCAACGT 592
    |||||
QY 2655 -GTAACCAACCGGCAAGACCAACATTAATTAAGTAGTACCGCTTGAGATTGAAT 2712
    |||||
DB 591 GAGCAATGAAGAGGACACCATGCTGCAACAGTAGACGTTGAGTGTGAGACTCTGCAC 532
    |||||
QY 2713 CAGTCCGTTGGCCTTAACCATCATCATCTCTGCAATTTGGGTGGAGCAA 2765
    |||||
DB 531 CAAGACGCTACACACACCCGATCATCAGCCCAACAAAGGTTGACATTA 479
    |||||

```

Search completed: July 31, 2004, 19:10:09  
Job time : 903 secs

This Page Blank (unintentional)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2004, 12:17:10 ; Search time 4825 Seconds  
(without alignments)  
18003.955 Million cell updates/sec

Title: US-09-700-712a-1  
Perfect score: 2909  
Sequence: 1 gagagagcttctatgcctc.....cagcgaatcgcgctgcag 2909

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues  
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: em\_escba:\*  
2: em\_escba:\*  
3: em\_escba:\*  
4: em\_escba:\*  
5: em\_escba:\*  
6: em\_escba:\*  
7: em\_escba:\*  
8: em\_escba:\*  
9: gb\_esc1:\*  
10: gb\_esc2:\*  
11: gb\_esc3:\*  
12: gb\_esc4:\*  
13: gb\_esc5:\*  
14: gb\_esc6:\*  
15: em\_escba:\*  
16: em\_escba:\*  
17: em\_escba:\*  
18: em\_escba:\*  
19: em\_escba:\*  
20: em\_escba:\*  
21: em\_escba:\*  
22: em\_escba:\*  
23: em\_escba:\*  
24: em\_escba:\*  
25: em\_escba:\*  
26: em\_escba:\*  
27: em\_escba:\*  
28: gb\_esc1:\*  
29: gb\_esc2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	249.8	8.6	1128	28	BZ577690 meh2_5528
2	194.8	6.7	805	28	BZ556310 pac81-60
3	191.8	6.6	853	28	BZ578932
4	161	5.5	1141	28	BZ559267

Result No.	Score	Query Match	Length	ID	Description
5	152	5.2	1375	28	BZ553447
6	115.6	4.0	406	29	U83227
7	111.2	3.8	1034	28	BZ561697
8	111	3.8	878	14	CD379073
9	111	3.8	921	14	CD382942
10	106.2	3.7	855	14	CD376447
11	104	3.6	843	14	CA588286
12	103.8	3.6	813	14	CA588344
13	103.2	3.5	828	14	CA475083
14	96	3.3	685	13	BM039803
15	96	3.3	686	13	BM039194
16	96	3.3	801	14	CA588156
17	92.4	3.2	662	14	CA588861
18	89	3.1	522	13	BM061136
19	89	3.1	632	13	BM295886
20	88.8	3.1	524	14	CD336403
21	86.6	3.0	591	12	BM039194
22	85	2.9	683	10	AM423104
23	84.8	2.9	700	14	CA531127
24	84.8	2.9	712	14	CA373071
25	84.6	2.9	483	29	CG637075
26	84.6	2.9	606	10	AM546106
27	84.6	2.9	668	13	BY746793
28	84.6	2.9	976	13	BY746114
29	84.6	2.9	2255	11	AK088606
30	84.6	2.9	2256	11	AK088134
31	84.6	2.9	3128	11	AK035663
32	84	2.9	677	14	CH420463
33	83.8	2.9	756	12	BI078307
34	82.8	2.8	776	14	CF995901
35	82.4	2.8	528	14	CD330233
36	82	2.8	748	13	BU055160
37	81.8	2.8	673	14	CD313851
38	81.4	2.8	407	14	CD330290
39	81.4	2.8	1121	11	AK011435
40	80.8	2.8	945	11	BF783860
41	80.8	2.8	2486	11	AK004832
42	80.2	2.8	952	10	BF782519
43	80.2	2.8	1024	12	BM924499
44	79.8	2.7	974	10	BF607905
45	79.2	2.7	567	9	AM108416

## ALIGNMENTS

RESULT 1  
BZ577690  
LOCUS  
DEFINITION meh2\_5528.y2 meh Pseudomonas aeruginosa genomic clone meh2\_5528,  
genomic survey sequence.

ACCESSION BZ577690  
VERSION BZ577690.1 GI:27212751

KEYWORDS  
SOURCE  
ORGANISM

Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.

REFERENCE  
1 (bases 1 to 1128)  
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,  
Burns,J.L., Kaul,R. and Oleen,M.V.

TITLE  
Whole-Genome-Sequence variation among multiple isolates of  
Pseudomonas aeruginosa library

JOURNAL  
U. Bacteriol. (2002) In press

COMMENT  
Contact: Chris K. Raymond  
Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA

Box 352145, Seattle, WA 98105-2145, USA  
Tel: 20622216954  
Fax: 2066857244  
Email: ckraymond@u.washington.edu

Classes: shotgun.  
Location/Qualifiers

FEATURES



```

RESULT 3
BZ578932      853 bp   DNA      linear   GSS 17-DEC-2002
LOCUS         msb_6070.v2  msb Pseudomonas aeruginosa genomic clone msb_6070,
DEFINITION    genomic survey sequence.
ACCESSION     BZ578932
VERSION       BZ578932.1  GI:27213993
KEYWORDS      GSS.
SOURCE        Pseudomonas aeruginosa
ORGANISM      Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE     1 (bases 1 to 853)
AUTHORS       Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Haastings,M.,
              Burns,J.L., Kaul,R. and Olsen,M.V.
              Whole-genome-sequence variation among multiple isolates of
              Pseudomonas aeruginosa library
              J. Bacteriol. (2002) in press
              Contact: Chris K. Raymond
              Genome Center
              University of Washington
              Box 352145, Seattle, WA 98105-2145, USA
              Tel: 2062216954
              Fax: 2066857244
              Email: craymond@u.washington.edu
              Class: shotgun.

FEATURES
  source
    1..853
    /organism="Pseudomonas aeruginosa"
    /mol_type="genomic DNA"
    /strain="MSH"
    /db_xref="taxon:287"
    /clone_1fb="msb_6070"
    /clone_1lb="msb"
    /note="Environmental isolate. Whole genomic shotgun
    library."

ORIGIN
  Query Match      6.6%; Score 191.8; DB 28; Length 853;
  Best Local Similarity 60.0%; Pred. No. 2e-44;
  Matches 336; Conservative 0; Mismatches 223; Indels 1; Gaps 1;

QY 29 CTGCAATTTCCCAATATGACCCCGATTTGTTTCATGCGCCCTTAGCGGCGCTG 88
DB 168 CTGACGTATCCCAATGATCGATCGGTTGCGCTGCGCCATCGCCCGCTGAATCACTG 227
QY 89 TATGCTGTATGATTTGGTGGGTTTCTTTTGTATGTGTGCGCAATCGCCGAGC 148
DB 228 TAGGGCTATGATGACTGATCGGGATCGGGGCGCTGCTGCTGCGCGCGCGATG 287
QY 149 GATCGCGCGGAGTGTGAGCGCGTGAAGCAATCTTGAATTTTTCGCGGCTTT 208
DB 288 AAGCGCTTGCAGCCGACCTGACCAAGAGCGCTTTCGACCTGCTTTCGGGTGCGC 347
QY 209 TTAGGTATGATGATGCGTGGCGAGTTGGTATGTATCTTCACAATTTTATCTGTC 268
DB 348 TCGGGGTATCTCGGTGCGCGCTGCGCTGAGTCTGTTCAACCTGACGAGTAC 407
QY 269 CTGTGACCTCTTATTTATTTCAAGTGTGAGCTGCGGCGCATGCTTCCACGCGCG 328
DB 408 ATCGCAACCGACGCGTATCTTCAGGTCTGGAAGGCGGCGATGCTTTCATGCGCG 467
QY 329 TTATGGGTGTATACCGCCCATGTTCTGGTATGCGCGTAAACCAACGACCTTTCTT 388
DB 468 CTGCTCGCGTATGCTGCGCGGTCTGTGTGCTGCGCAAGCGCCATGCAAGCTTCTC 527
QY 389 GGTGTGCGGATTTGTGCGCCCTTAGTGCATTTGGTTGGGATGAGATGATCGAT 448
DB 528 CAGCTGATGATCTTATCGCCCGCTGTGCGCCATGCGCGCGCGCGCGCATCGCG 587
QY 449 AACTTATGATGATGATCTTGGGAGCAAGTAAACGATGATGCTTGGCTTTGTATTC 508
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```

DB 588 AACTTCATCACTCGAACTGTGGGCAAGGTCAACGATGTCCTTGGCCATGCTTTC 647
QY 509 CCTATGTGCGCCCACTGCGCGGCATCTTCAAGCTTTATGAATTCGCTTAGAGGC 568
DB 648 CCCAAACG-CGCGCGCTGCGCGCATCTCTGCACTGTACCAAGTCCCTTGAAAGG 706
QY 569 GTGTCTGTTCTTTATCT 588
DB 707 CGTGGCGTGTCTCATCTCT 726

RESULT 4
BZ559267      1141 bp   DNA      linear   GSS 17-DEC-2002
LOCUS         pac82-164_1355.s1 pac82-164 Pseudomonas aeruginosa genomic clone
DEFINITION    pac82-164_1355, genomic survey sequence.
ACCESSION     BZ559267
VERSION       BZ559267.1  GI:27175318
KEYWORDS      GSS.
SOURCE        Pseudomonas aeruginosa
ORGANISM      Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE     1 (bases 1 to 1141)
AUTHORS       Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Haastings,M.,
              Burns,J.L., Kaul,R. and Olsen,M.V.
              Whole-genome-sequence variation among multiple isolates of
              Pseudomonas aeruginosa library
              J. Bacteriol. (2002) in press
              Contact: Chris K. Raymond
              Genome Center
              University of Washington
              Box 352145, Seattle, WA 98105-2145, USA
              Tel: 2062216954
              Fax: 2066857244
              Email: craymond@u.washington.edu
              Class: shotgun.

FEATURES
  source
    1..1141
    /organism="Pseudomonas aeruginosa"
    /mol_type="genomic DNA"
    /strain="2-164"
    /db_xref="taxon:287"
    /clone_1fb="pac82-164_1355"
    /clone_1lb="pac82-164"
    /note="clinical isolate 2-164 Whole genomic shotgun
    library."

ORIGIN
  Query Match      5.5%; Score 161; DB 28; Length 1141;
  Best Local Similarity 64.6%; Pred. No. 2.7e-35;
  Matches 272; Conservative 0; Mismatches 145; Indels 4; Gaps 2;

QY 380 ACCCTCTTGTGTGCGCGATTTTGTGCGCCCTTAGTGCATTCGTTTGGGATGGA 439
DB 628 ACTCTTCCACACTATGACTTCAATCGCCCGCTGTGCGCATCGGCTGGCGCCGCG 569
QY 440 CGTATCGTACTTATGATGATGAACTTTGGGACAGTAAACGATGTCCTTGGCTT 499
DB 568 CGCATCGGCAACTTATCAACTCGGAACGTGGGGCAAGTCAAGATGTGCTTGGGCG 509
QY 500 TTGTATTCCTCAATAGTGGCGCCACGTGCGCGGCATTCCTCAACGCTTATGAATTC 559
DB 508 ATGTCTTCCCAACGCGCGAC-C-GTGGCGGCAATCCTCGACCTGTACAGTTGCGC 450
QY 560 TTAGAAGCGTGTCTGTCTTTATTTCTTATTTATTTATTTATTTATTTATTTAT 619
DB 449 CTGAAGGAGTGTGGCGCTGTGTCTATCTCTGCTTCAACCGCAAGCAAGCGCGACC 390
QY 620 GCGACGATTCGGAAGCTTTTATCTGATACGATACGATTCCTTCTTGTGAATAC 679
DB 389 GCTGTCTTCCGCGCTGTGTGCTGTGCTACGGAATCTTCCGCTTGTTGTGAATTC 330
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

QY 680 GTCCGTGACCGAGATGCTCAGTTGGG--TCTGTTGGTGCTTCATTCAATGGGGCA 736  
 Db 329 GTCCCGTGGCCGGATGACCCAGCTCGCTGGGGTTGGCTGACCATGGGCGCA 270  
 QY 737 ATCTCTCTCTTACCTATGATGATCATCGGATTTTGGATGATGATGTTGGCTTACAGCGC 796  
 Db 269 GTGCTCTGTGATCCAGATGCTGCGCGGATTCCTGATGATGCTGCGGCTTACGCGCGC 210  
 QY 797 G 797  
 Db 209 G 209

RESULT 5  
 BZ53447 1375 bp DNA linear GSS 17-DEC-2002  
 LOCUS pacel-60\_4149, genomic survey sequence.  
 DEFINITION BZ53447  
 VERSION BZ53447  
 KEYWORDS BZ53447.1 GI:27159498  
 SOURCE GSS.  
 ORGANISM Pseudomonas aeruginosa  
 Pseudomonas aeruginosa  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Pseudomonadaceae; Pseudomonas.  
 REFERENCE 1 (bases 1 to 1375)  
 Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,  
 Burns, J.L., Kaul, R. and Olsen, M.V.  
 Whole-genome-sequence variation among multiple isolates of  
 Pseudomonas aeruginosa library  
 J. Bacteriol. (2002) in press  
 CONTACT: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 20622216954  
 Fax: 2066857244  
 Email: craymond@u.washington.edu  
 Class: shotgun.  
 FEATURES  
 source location/Qualifiers  
 1..1375  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="1-60"  
 /db\_xref="taxon:287"  
 /clone="pacel-60\_4149"  
 /clone\_lib="pacel-60"  
 /note="Clinical isolate 1-60 Whole genomic shotgun  
 library."

ORIGIN  
 Query Match 5.2%; Score 152; DB 28; Length 1375;  
 Best Local Similarity 57.7%; Pred. No. 1.3e-32;  
 Matches 291; Conservative 0; Mismatches 210; Indels 3; Gaps 1;  
 270 TTGCTGACCTCTTATTTATTCACAAAGTGTGACGTGCGGCATGCTCTTCCAGCGGCT 329  
 |||||  
 84 TTCCTGACGCCCCGCGATTTCTAGGTCTGCAAGGGCGGATGCTCTTCCATGCGGGC 143  
 |||||  
 330 TATTGGGTGTATCACCGGCATGTTCTGTATGCGGATGAAAACCAAGCAGCTTTCTTGG 389  
 |||||  
 144 TGCTGGCGGTATGCTGCGGTCTGTGTGTTGCGCAAGCGCCATGCAAGAGCTTCTTC 203  
 |||||  
 390 GGTGCGCGATTTTGTGTCCTTTAGTGCCATTTGGGTTTGGAGATGGAGCGTATCGGTA 449  
 |||||  
 204 AGCTATGAGACTTATCGCCCGCTGCTGTCCTGAGCTGAGCGCGCGGCGCATCGCA 263  
 |||||  
 450 ACTTATGATATGTAACCTTTGGGACGAGTAACGAGATGCTGCTTGGCTTTTGTATTC 509  
 |||||  
 264 ACTTATCAACTCGGAAGTGTGGGGAAGGCGCATATGAGCCCTGGGCGCATGCTTCC 323  
 |||||  
 510 CTAATGTGGCCCATGCGCGCATCTTCAAGCTTTATGAATTCGCTTGAAGGCG 569  
 |||||

Db 324 CCAAGCGGCGCCGCTGCGCGGACATCCCTGAAAGTTGATCAAGTTGCGCCCTGAAAGGCC 383  
 QY 570 TGTCTTCTTCTTATTTCTTATTTATGTTTATGTTAACTGTCGCCGTAGGACGAT 629  
 |||||  
 Db 384 TGGACATTTGTGATCTCTGCGTGGCAACCCGACGACGATGACCGCTTGTCT 443  
 |||||  
 QY 630 CCGACGTGTTTATGCTGGAATACGATATTCGCTTCTTGTGGAATAGTCCGTGAGC 689  
 |||||  
 Db 444 CTGGCTGTGTGCTGTGCTACCGGATCTACCGCTTGTGATGGGGAATTCACCGCTGTC 503  
 |||||  
 QY 690 CAGATGCTAGTTGG--GTCGTTTGTGCTTCATTTCAATGGGGAATCCCTCTCT 746  
 |||||  
 Db 504 CGATGCCACCTGCTTACTTCCATGAGTTGTGCTGACCATGGGCCAAGAGCTGCCAG 563  
 |||||  
 QY 747 TACCTATGATCATCGATATT 770  
 |||||  
 Db 564 TACCGTGGAAATGCCCTTCT 587  
 |||||

RESULT 6  
 U83227 406 bp DNA linear GSS 21-FEB-2001  
 LOCUS U83227  
 DEFINITION U83227 Pseudomonas putida Mmb1: ATCC 23483 Pseudomonas putida  
 genomic clone transposon-tagged mutant UT502 similar to  
 pilIpoprotein diacylglycerol transferase (19c), genomic survey  
 sequence.  
 REFERENCE U83227  
 U83227.1 GI:1786113  
 SOURCE GSS.  
 ORGANISM Pseudomonas putida  
 Pseudomonas putida  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Pseudomonadaceae; Pseudomonas.  
 REFERENCE 1 (bases 1 to 406)  
 Caspi, R., Tebo, B.M. and Haygood, M.G.  
 c-type cytochromes and manganese oxidation in Pseudomonas putida  
 Mmb1  
 Appl. Environ. Microbiol. 64 (10), 3549-3555 (1998)  
 JOURNAL 98432765  
 MEDLINE 9758766  
 PUBMED  
 CONTACT: Caspi R  
 MBRD  
 Scripps Institution of Oceanography, UCSD  
 9500 Gilman Drive, La Jolla, CA 92093-0202, USA  
 Email: rcaspi@ucsd.edu  
 This sequence was obtained by automated sequencing from one strand  
 of the DNA. It was NOT confirmed by sequencing the second strand.  
 Therefore, some mistakes might be present.  
 Class: transposon-tagged.  
 FEATURES  
 source location/Qualifiers  
 1..406  
 /organism="Pseudomonas putida"  
 /mol\_type="genomic DNA"  
 /strain="Mmb1; ATCC 23483"  
 /db\_xref="taxon:303"  
 /clone="transposon-tagged mutant UT502"  
 /clone\_lib="Pseudomonas putida Mmb1; ATCC 23483"  
 /note="manganese oxidizing bacterium"

ORIGIN  
 Query Match 4.0%; Score 115.6; DB 29; Length 406;  
 Best Local Similarity 59.8%; Pred. No. 3.8e-22;  
 Matches 234; Conservative 0; Mismatches 145; Indels 12; Gaps 2;  
 399 ATTTTGTGCCCCCTTATGATGCAATTCGTTGGGATGAGAGATGCGTAACCTTATGA 458  
 |||||  
 15 ATTTCGTGCCCCCGCTGTGTCGATCGGCTGCGGCGGCGATCGGCAATTCATCA 74  
 |||||  
 459 ATAGGAATTTGGGACGAGTAACGAGATGCTTGGGCTTTTGTATTTCC----- 510  
 |||||  
 75 ACGCGAGCTGTGGGCAAAACCAACGACGTGCGCATGATCTTCCCGCGCTTCA 134  
 |||||  
 511 -TAATGTGGCCCATGCGCGCATCTTCAAGCTTTATGAATTCGCTTGAAGGCG 569  
 |||||

Db 135 GGCACCCAGCCAGCTGCCAGCCAGCTGATACAGTTGCGCCGTGAGGGCG 194  
 QY 570 TGGTTCTGTTCTTATTTTAAATGTTATTTGTAACCTGTCGGTAGGAGGTAAT 629  
 Db 195 TGGCACTGTTGTCATCTCTGCTGCTTCTCGCGAAGCGCGCGGAGCCAGTGG 254  
 QY 630 CCGGACGTGTTTATGCTGATAGATACATTCGCTTCTTGTGAAATAGTCCTGAGC 689  
 Db 255 CCGGACGTGTTGCGGCTGTTCTAGGCAATCTCGCTTCACTGTCGAATTCGCGGTGC 314  
 QY 690 CAGATGCTCAGTTGGG---TCTGTTTGTGGCTTCAATTCATGAGGCAATCTCTCT 746  
 Db 315 CGGATGCCAGCTGGGCTATATGCGCTGGGGCTGGCTGACATGGGTGAGATTCTGTGCG 374  
 QY 747 TACCTATGCTGATCATCGGTATTTTATGAT 777  
 Db 375 TGCCGATGATCTGCTGCTGGCTTGGCGCTGAT 405

## RESULT 7

LOCUS BZ561697 1034 bp DNA linear GSS 17-DEC-2002  
 DEFINITION pac82-164\_3389.y2 pac82-164 Pseudomonas aeruginosa genomic clone  
 ACCESSION BZ561697  
 VERSION BZ561697.1 GI:27182207  
 KEYWORDS GSS.

SOURCE Pseudomonas aeruginosa  
 ORGANISM Pseudomonas aeruginosa  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.

REFERENCE 1 (bases 1 to 1034)  
 Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Haerings, M., Burns, J.L., Kaul, R. and Olsen, M.V.  
 TITLE Whole-Genome-Sequence Variation among Multiple Isolates of Pseudomonas aeruginosa Library  
 J. Bacteriol. (2002) In Press  
 JOURNAL  
 COMMENT Contact: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216954  
 Fax: 2066857244  
 Email: craymond@u.washington.edu  
 Classes: Shotgun.

FEATURES  
 source Location/Qualifiers  
 1..1034  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="2-164"  
 /db\_xref="taxon:287"  
 /clone="pac82-164\_3389"  
 /clone\_1b="pac82-164"  
 /note="clinical isolate 2-164 Whole genomic shotgun library."

## ORIGIN

Query Match 3.8%; Score 111.2; DB 28; Length 1034;  
 Best Local Similarity 53.9%; Pred. No. 1.2e-20;  
 Matches 342; Conservative 0; Mismatches 278; Indels 14; Gaps 5;  
 QY 29 CTGCAATTTCCCAATTTGACCCCGATTTGTTTGCATGCGCCCTTAGCGGTCCTGG 88  
 Db 370 CTGACCTATCCCGAGATCATCGATGCGTGGCCGATCGGCCGTGAAGATCCACTGG 429  
 QY 89 TATGCTGATGATTTGTTGTTGTTCTTTTGTATGTTGTTGTTGTTGTTGTTGTTGTT 148  
 Db 430 TACGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 489  
 QY 149 GATCGCGCGCGGAGTGTGAGCGCGTGAAGCAATCTTGAATTTGTTATTC---GCCGCG 205  
 Db 490 AAGCGCTTGACCGGACCGATGACCAAGAGCGCGCTTTCGACCTGCTTCTGCTGGTTCCG 549

QY 206 TTTTATGTTATGATATCGATGCGGAGTTGGTTATGATGATCTTACATTTTATCTG 265  
 Db 550 CTGGGCGCGTATTCCTCGGTGGCGCCCTGGGCTAGTCTTCTTACAACTGAGCGAG 609  
 QY 266 TTCTTGTGATCCCTCTTTATTTATTTAAAGTTGATGATGCGGCGATGCTTTCAGCGC 325  
 Db 610 TACATGCGCAACCGACGCTGATCTTCAGAGTCTGGGAAGGGCGGAGTCTTTCATGCG 669  
 QY 326 GCGTATTTGGGTGATATCACCGCCATGTTCTGTATGCGCGTAAACCAAGCACCTTC 385  
 Db 670 GGGCTGCTCGGCTGATGCTGCGGCTGCTGTTGTTTGGCAGGCGCCATGCGAAGAGC 729  
 QY 386 TTTGGTGTGGCGA-----TTTGTGCCCCCTTATGATGCAATTCGTTGGGATGAGAC 440  
 Db 730 TTCTTACGCTGAGGAGATTTGATTCGCCCCCTTGTGTCATGAGCCCTGGCGCGCGCC 789  
 QY 441 GATCGTAACTTTATGAAATAGTGAACCTTGGGAGAGAG--TAAAGATGTCCTTGG 497  
 Db 790 GCATTTGGAAACTTTTATTAATGTAACCTTGTGGCGAAGGTCAATGATGAGCCCGGCG 849  
 QY 498 CTTTGTATTCCTTAATGTTGCGCCCACTGCGCGCCCATCTTACAGCTTATGAA-TTC 556  
 Db 850 CAGGCGCTTTCCTCCCAACGTGCGCCCTTTCGCGGCGATCCCTTACATTTGCCAATTC 909  
 QY 557 GCCTAGAGAGCGTGTCTGTTCTTTA--TTCTTAATGTTTATGTTAAACCTGTC 614  
 Db 910 GCTTGAAGGCGGCGGCGCGCTTTTATATCTTGTGTTAGTTAACCCCAAGCCGCGCG 969  
 QY 615 CGCTAGGACGCTATCCGAGCTGTTTATGACTGG 648  
 Db 970 CGCGCGCCTTGTTTCCGTCTTTCGCGCTTG 1003

## RESULT 8

LOCUS CD379073/c 878 bp mRNA linear EST 31-MAY-2003  
 DEFINITION PTM004242 Phaeodactylum tricornutum Uni-Zap XR Phaeodactylum tricornutum cDNA 5', mRNA sequence.

ACCESSION CD379073  
 VERSION CD379073.1 GI:31254687  
 KEYWORDS EST.  
 SOURCE Phaeodactylum tricornutum  
 ORGANISM Phaeodactylum tricornutum  
 Eukaryota; stramenopiles; Bacillariophyta; Bacillariophyceae; Bacillariophycidae; Naviculales; Phaeodactylaceae; Phaeodactylum.

REFERENCE 1 (bases 1 to 878)  
 Scala, S., Carels, N., Falciatore, A., Chiusano, M.L. and Bowler, C.  
 TITLE Genome properties of the diatom Phaeodactylum tricornutum  
 JOURNAL Plant Physiol. 129 (3), 993-1002 (2002)  
 MEDLINE 22111123  
 PubMed 12114555  
 COMMENT Contact: Bowler C  
 Laboratory of Molecular Plant Biology  
 Stazione Zoologica 'Anton Dohrn'  
 Villa Comunale, I-80121, Napoli, Italy  
 Tel: 39 081 583 3268/3211  
 Fax: 39 081 764 1355  
 Email: chris@alpha.szn.it  
 Diatom EST Database (http://aves.thegen.szbowler.com)  
 Seq primer: T3 backward  
 POLYA=yes.

FEATURES  
 source Location/Qualifiers  
 1..878  
 /organism="Phaeodactylum tricornutum"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:2850"  
 /cell\_line="CMP632"  
 /clone\_1b="Phaeodactylum tricornutum Uni-Zap XR"  
 /note="Vector: Uni-Zap XR vector; Site\_1: Eco RI; Site\_2: Xho I"

## ORIGIN

Xho I"

	Query Match	Similarity	59.8%	Score	111;	DB	14;	Length	878;	
	Best Local	Similarity	59.8%	Pred.	No. 1,2e-20;					
	Matches	186;	Conservative	0;	Mismatches	125;	Indels	0;	Gaps	0;
OY	2474	ATCATTTCTTAGAGGTAGGAAAGATCGACCCGGAGAGAATTGAAAAAATTCGTGATGSGT	2533							
Dp	332	ATGACTTCTTAACAGGAAGAAGTATGGCCACCAGCACAAGAAATTAAAATCATATCGTGACGGSTG	273							
OY	2534	GCACCTGGGGAAGACCGCGGAACCTCTTTTACAGCCCATATAGGCCCAAGGCTGACCGGA	2593							
Dp	272	GCACCAAGGAAAGCACCGCTCGAGCTGATGCCCATGCCCCCATTTGGCCCATAGCGACGATT	213							
OY	2594	GTATTGTGCACAGTAGTACCAATATTGGCACCCCATCACATAGAAATGCGGGTTTCAAC	2653							
Dp	212	GTGTGTGTGATCTGAAGTACCAATATTCCTCTCCCATGATCATGTAAATTCCTTGGGGACG	153							
OY	2654	GATAACCCACCGGCAACGACGACCAACAATATAGAAAGTACCGTCTTAGAGATTGAAT	2713							
Dp	152	GAGACAGCGGATTCGACCAAGAACACACAACAATCGAGGTCTGTGTAGAAAGAACTGAGC	93							
OY	2714	AGTGCCTGTCGACCTAAACCAATCATCATCTCTGCAATTTGGTGGGAAGCAAATTCAAAT	2773							
Dp	92	AATCAGTCGCAAGAAATACCAACCATTTAGACGACGAAATGGGATTCGTCTGTCGCGCAAC	33							
OY	2774	AGACTTTGGC	2784							
Dp	32	AACCTCCCGC	22							
RESULT 9	CD382942/c									
LOCUS	PTM08112	Phaeodactylum tricornutum Uni-Zap XR Phaeodactylum								
DEFINITION	tricornutum cDNA 5', mRNA sequence.									
ACCESSION	CD382942									
VERSION	CD382942.1									
KEYWORDS	EST.									
SOURCE	Phaeodactylum tricornutum									
ORGANISM	Phaeodactylum tricornutum									
	Bukaryota; Eremophilales; Bacillariophyta; Bacillariophyceae;									
	Bacillariophycidae; Naviculales; Phaeodactylaceae; Phaeodactylum.									
REFERENCE	1 (bases 1 to 921)									
AUTHORS	Scala,S., Carels,N., Falciatore,A., Chisano,M.L. and Bowler,C.									
TITLE	Genome properties of the diatom Phaeodactylum tricornutum									
JOURNAL	Plant Physiol.									
MEDLINE	22111123.									
PUBMED	12114555									
COMMENT	Contact: Bowler C Laboratory of Molecular Plant Biology Stazione Zoologica 'Anton Dohrn' Villa Comunale, I-80121, Napoli, Italy Tel: 39 081 583 3268/3211 Fax: 39 081 764 1355 Email: chris@alpha.szn.it Diatom EST Database( <a href="http://aves.thagen.szbw.biology.com">http://aves.thagen.szbw.biology.com</a> ) Seq primer: T3 backward POLYA=yes.									
FEATURES	source									
	location/Qualifiers									
	1..921									
	/organism="Phaeodactylum tricornutum"									
	/mol_type="mRNA"									
	/db_xref="caxon:2850"									
	/cell_line="CCMP632"									
	/clone_lib="Phaeodactylum tricornutum Uni-Zap XR"									
	/note="Vector: Uni-Zap XR vector; Site_1: Eco RI; Site_2: Xho I"									
ORIGIN										
	Query Match	3.8%;	Score	111;	DB	14;	Length	921;		
	Best Local	Similarity	59.8%;	Pred.	No. 1.3e-20;					
	Matches	186;	Conservative	0;	Mismatches	125;	Indels	0;	Gaps	0;
OY	2474	ATCATTTCTTAGAGGTAGGAAAGATCGACCCGGAGAGAATTGAAAAAATTCGTGATGSGT	2533							

Dd		353	ATGACTTCTACAGGAAGAAGTAATGGCCACCGACAAGAAATTAAACATATCGTGGA CGGTG	294
Oy		2534	GCACTGGCGAAGAACA CGGCCAAATCTTTCTTTAACAGGCATATGGCCAAAGCTGACGAGA	2593
Dd		293	GCACACAGCGAAGAAGCAGCTCGACGTGATCCCATTCGCCCATTTGGCCCATGACGACATT	234
Oy		2594	GTATTGGTCACAGTAGTACCAAATTATGGCAACCATCACCATAGGAATGGCGTTTCAAC	2653
Dd		233	GTGTTGGTGCATGAAGTACCAATATTTGGCTTCATGATCATGTAAATTTCTTGGCGGACG	174
Oy		2654	GGTAAACCCACCGCGAACAAGACCAACAATAATAGAAGTCAACCGTCTTGAGATTTGAATC	2713
Dd		173	GAGACAGCGATTCGACCAAGAACAACAATCGAGGTCGTTTGAAGAAGAAGACTGAGGC	114
Oy		2714	AGTGCCTGTGGCACTAAACCAATATCATCATCTTGCAATTTGGGTGGGAAGCAATTCAAT	2773
Dd		113	AATACAGTCGCAAGAAATACCAACCATTAAGACAGCAATGGGATTCGTGTCGCGCAAC	54
Oy		2774	AGAACTTTGGC 2784	
Dd		53	AAC TCTCGGC 43	
RESULT 10				
LOCUS	CD376447/c	855 bp	mRNA linear	EST 31-MAY-2003
DEFINITION	PTMM01614 Phaeodactylum tricornutum Uni-Zap XR Phaeodactylum			
ACCESSION	CD376447			
VERSION	CD376447.1	GI:31252061		
KEYWORDS	EST.			
SOURCE	Phaeodactylum tricornutum			
ORGANISM	Phaeodactylum tricornutum			
REFERENCE	Eukaryota; Euxenopiles; Bacillariophyta; Bacillariophyceae;			
AUTHORS	Bacillariophycidae; Naviculatales; Phaeodactylaceae; Phaeodactylum.			
TITLE	1 (bases 1 to 855)			
JOURNAL	Scala,S., Carrels,N., Falciatore,A., Chiusano,M.L. and Bowler,C.			
MEDLINE	Genome properties of the diatom Phaeodactylum tricornutum			
PUBMED	Plant Physiol. 129 (3), 993-1002 (2002)			
COMMENT	12114555			
	Contact: Bowler C			
	Laboratory of Molecular Plant Biology			
	Stazione Zoologica 'Anton Dohrn'			
	Villa Comunale, I-80121, Napoli, Italy			
	Tel: 39 081 583 3268/3211			
	Fax: 39 081 764 1355			
	Email: chris@alpha.szn.it			
	Diatom EST Database http://aves.chagen.sznbowler.com)			
	Seq Primer: T3 backward			
	POLYA=Yes.			
FEATURES				
source	location/Qualifiers			
	1..855			
	/organism="Phaeodactylum tricornutum"			
	/mol_type="mRNA"			
	/db_xref="taxon:2850"			
	/cell_line="CCMP632"			
	/clone_id="Phaeodactylum tricornutum Uni-Zap XR"			
	/note="Vector: Uni-Zap XR vector; Site_1: Eco RI; Site_2:			
	Xho I"			
ORIGIN				
Query Match	3.7%; Score 106.2; DB 14; Length 855;			
Best Local Similarity	58.8%; Pred. No. 3.1e-19;			
Matches 183; Conservative	0; Mismatches 128; Indels 0; Gaps 0;			
Oy	2474 ATCAATTTCTAGAGGTAGAGACATCAACAACCGCGAAGAGATTGAAAATAATCGTGAATGATG	2533		
Dd	323 ATGACTTCTACAGGAAGAAGTAATGGCCACCGACAAGAAATTAAACATATCGTGACGGTG	264		
Oy	2534 GCACTGGCGAAGAAGCAGGAATCTTTCTTTTACAGGCATATGGCCAAAGCTGACGAGA	2593		

Db 263 GCACGACGGAAGACGCTGAGTGTGATCCCATGCGCCATTTGGCCCATAGCAGATG 204  
 QY 2594 GTATTGTCACAGTATGATCAATATTTGGACCATCATCAGCAATTCGGTTCAACC 2653  
 Db 203 GTGTTGGTACAGTACAGTACCAATTTTCCTCCCATATCATGTAAATACCTTGGCGGAG 144  
 QY 2654 GGTAAACCAACCGGCAACGAGCAACCAATATGAAATGACCGTGTGAGATTGAATC 2713  
 Db 143 GAGACAGCGGATCCGACCAAGACCAACATCGAGTATGTTGTAAGACGATCGAAGC 84  
 QY 2714 AGTCCGTTTCCATCAATCAATCATCTCTGCAATTTGGTGGAGCAATTTCAAT 2773  
 Db 83 AATACAGTCGCAAGATACCAACATTTAGACGAGATGATGCTGTCTGCGCGAAGC 24  
 QY 2774 AGACTTTGGC 2784  
 Db 23 AACTCTCCGCGC 13

RESULT 11 843 bp mRNA linear EST 19-NOV-2002  
 CAS88286/c  
 LOCUS hab9d11.y1 Fugu UT6 adult gut Takifugu rubripes cDNA clone  
 IMAGE:6351812 5' similar to TR:Q91237 Q91237 NA/PI CONTRAST  
 SYSTEM PROTEIN. ; mRNA sequence.

ACCESSION CAS88286  
 VERSION CAS88286.1 GI:25132864  
 KEYWORDS EST  
 SOURCE Takifugu rubripes (Fugu rubripes)  
 ORGANISM Takifugu rubripes  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 Acanthomorphi; Acanthopterygii; Perciformes; Tetraodontiformes;  
 Tetraodontidae; Tetraodontidae; Takifugu.  
 1 (bases 1 to 843)  
 Clark, M., Johnson, S. L., Lehrach, H., Lee, R., Li, F., Marra, M.,  
 Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,  
 Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,  
 Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk, R.,  
 Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,  
 Waterston, R. and Wilson, R.  
 WashU Zebrafish EST Project 1998  
 Unpublished (1998)  
 Contact: Stephen L. Johnson  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: zbrafish@wustl.edu  
 Library materials provided by G. Elgar (UK MRC HGMP-RC) Library  
 constructed by Drs. K. Kawakami, M. Sasaki, S. Sugano, K. Kikuchi  
 and S. Watabe (University of Tokyo, Institute of Medical Science  
 and Laboratory of Aquatic Molecular Biology and Biotechnology) DNA  
 Sequencing by: Washington University Genome Sequencing Center Clone  
 distribution: Fugu clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LIML, send email to:  
 info@image.liml.gov  
 Seq primer: T3 Et from Amersham  
 High quality sequence stop: 496.  
 Location/Qualifiers  
 1..843  
 /organism="Takifugu rubripes"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:31033"  
 /clone="IMAGE:6351812"  
 /cdate="1998-11-19"  
 /dev\_stage="adult"  
 /lab\_host="PH108 (T1 phage-resistant)"  
 /clone\_lib="Fugu UT6 adult gut"  
 /note="Vector: pBluescript-Pu; Site 1: PflM I  
 (CCANNNNNTGG); Site 2: PflM I (CCANNNNNTGG); BamHI-SmaI  
 sites were converted to BamHI-PflM1-SfiI-PflM1 sites (SmaI  
 is destroyed). Other part of the vector is untouched. The

ORIGIN  
 Query Match 3 6%; Score 104; DB 14; Length 843;  
 Best Local Similarity 58.9%; Pred. No. 1.4e-18;  
 Matches 198; Conservative 0; Mismatches 135; Indels 3; Gaps 1;

CDNA is inserted between two PflM1 sites in T3 (5') to T7  
 (3') direction. Library materials provided by G. Elgar (UK  
 MRC HGMP-RC) and constructed and donated by Drs. K.  
 Kawakami, M. Sasaki, S. Sugano, K. Kikuchi and S. Watabe  
 (University of Tokyo, Institute of Medical Science and  
 Laboratory of Aquatic Molecular Biology and  
 Biotechnology)."

Db 2427 GCAGAGCGAAGACGACCGAGTGTGAGACTTTCTTAATGCGCAACATCATTTCTAGAG 2486  
 Db 721 GGAAGAGCTAATCAAGAACTGTGGAAGTGTAAAGAACTCTGTAGCAGCTCAGAG 662  
 QY 2487 GTAGAAAGTCAGACCGCGAGAGATGAAAAAATCGTGATGCTGCACTGGCGAAG 2546  
 Db 661 GCAGAAAGATCAACACGACGACGATTAAGAAAGTGTGAGAGGTCGCCAGCAAGG 602  
 QY 2547 CACGCGGAACTCTTTTACAGCGCATATGCGCAAGCTGACGAGATTTGTACAG 2606  
 Db 601 CCTTCGGAACCTGTTTGATCCCTCCGCTGATCATGCGCCAGATGTTGTGACAG 542  
 QY 2607 TAGTACCAATATTGACCCATCATCATAGATGAGATGCGGTTT---CAACCGTAACCCAC 2663  
 Db 541 AGTCCCATATGTGGCGCCCATGATTAATCGGACCGCGCTGTGACATCATGATTTCCAG 482  
 QY 2664 CGGACAGACGACCAACATATATGAAATGACCGTCTTGAGATGTAATGATGCGGTTG 2723  
 Db 481 AGGACACCATGTGTGACCAACATATGAGAGAGAGTCTGAGAGCTGTGACAGCAGGTGA 422  
 QY 2724 CCATTAACCATCATCATCTGCAATTTGGTGGG 2759  
 Db 421 CTAAACCCCGATCAACGACGACGACGAGGTTG 386

RESULT 12 813 bp mRNA linear EST 19-NOV-2002  
 CAS88344/c  
 LOCUS hab50c01.y1 Fugu UT6 adult gut Takifugu rubripes cDNA clone  
 IMAGE:6351745 5' similar to TR:Q91237 Q91237 NA/PI CONTRAST  
 SYSTEM PROTEIN. ; mRNA sequence.

ACCESSION CAS88344  
 VERSION CAS88344.1 GI:25132922  
 KEYWORDS EST  
 SOURCE Takifugu rubripes (Fugu rubripes)  
 ORGANISM Takifugu rubripes  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 Acanthomorphi; Acanthopterygii; Perciformes; Tetraodontiformes;  
 Tetraodontidae; Tetraodontidae; Takifugu.  
 1 (bases 1 to 813)  
 Clark, M., Johnson, S. L., Lehrach, H., Lee, R., Li, F., Marra, M.,  
 Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,  
 Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,  
 Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk, R.,  
 Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,  
 Waterston, R. and Wilson, R.  
 WashU Zebrafish EST Project 1998  
 Unpublished (1998)  
 Contact: Stephen L. Johnson  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: zbrafish@wustl.edu  
 Library materials provided by G. Elgar (UK MRC HGMP-RC) Library  
 constructed by Drs. K. Kawakami, M. Sasaki, S. Sugano, K. Kikuchi  
 and S. Watabe (University of Tokyo, Institute of Medical Science  
 and Laboratory of Aquatic Molecular Biology and Biotechnology) DNA  
 Sequencing by: Washington University Genome Sequencing Center Clone

distribution: Fugu clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov

Seq primer: T3 RT from Amersham  
High quality sequence stop: 497.

## FEATURES

source

1. 813  
Location/Qualifiers  
/organism="Takifugu rubripes"  
/mol\_type="mRNA"  
/db\_xref="taxon:31033"  
/clone="IMAGE:6351745"  
/issue\_type="gut"  
/dev\_stage="adult"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_id="Fugu UT6 adult gut"  
/note="Vector: pBluescript-PL; Site 1: PflM I (CCANNNNNTGG); Site 2: PflM I (CCANNNNNTGG); BamHI-SmaI sites were converted to BamHI-PflM-SfiI-PflM sites (SmaI is destroyed). Other part of the vector is untouched. The cDNA is inserted between two PflM sites in T3 (5') to T7 (3') direction. Library materials provided by G. Elgar (UK MRC HGP-RC) and constructed and donated by Drs. K. Kawakami, M. Sasaki, S. Sugano, K. Kikuchi and S. Watabe (University of Tokyo, Institute of Medical Science and Laboratory of Aquatic Molecular Biology and Biotechnology)."

## ORIGIN

Query Match 3.6%; Score 103.8; DB 14; Length 813;  
Best Local Similarity 59.2%; Pred. No. 1.6e-18;  
Matches 196; Conservative 0; Mismatches 132; Indels 3; Gaps 1;  
2432 GGGCAAGAGCCAGTGTGAGACTTCTCTAAATGCAACATCATTTCTAGAGTGG 2491  
713 GACGTAATCAGAAATGGGTGAACCTGTAAAGAACTCTGTAGCCACCTCCAGAGCAGA 654  
2492 AAGATCAGCAGCCGAGAAATGAAAAATCGTGAATGTGGCACTGCGCAAGACACGG 2551  
653 AGGATCAACACCGACAGCAAGTTAAAGAGTCGTGACGGTCGCCCAAGAAAGCCCTG 594  
2552 CGAAATCTTTCTTACAGCGCATATGGCCAGAGTGAAGATTTGGTCACTATTA 2611  
593 CGGAATCGTTTCGATCCCCCGCTGCATCATGCGACAGATGTTGTGTGACAAGATC 534  
2612 CCAATATTGGACCCATCATCATAGAAATCGCGGTTT---CAACCGGTAAACCCACGGCA 2668  
533 CCGATGTTGGCGCCCATGATATCGGCACCGCGCTGTGACATCCAGTAATTCAGAGAC 474  
2669 ACGAGACCAACAAATTAAGAAAGTCAAGTCTTGAAGATTGAATCAATGCGCTTGCAC 2728  
473 ACCATGTGACCAACATAGAGAGAGAGTGTGAGCTGTGACGAGACGCTGACTAAC 414  
2729 AAACCAATCATCATCTGCAATTGGGTGG 2759  
413 ACCCGATCACCAAGCCACGACGAGGTTGG 383

## RESULT 13

CA475083/c 828 bp mRNA linear EST 12-NOV-2002  
LOCUS CA475083  
DEFINITION AGENDCORT.10670196 NCI CGAP ZKId1 Danio rerio cDNA clone  
IMAGE:6797422 5', mRNA sequence.  
ACCESSION CA475083  
VERSION CA475083.1 GI:24931435  
KEYWORDS EST.  
ORGANISM Danio rerio (zebrafish)  
SOURCE Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes; Cyprinidae; Danio.  
REFERENCE 1 (bases 1 to 828)  
AUTHORS NIH-MGC <http://mgc.mci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

## JOURNAL

Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov

Tissue Procurement: Leonard I. Zon, M.D.  
cDNA Library Preparation: Invitrogen Corp.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Plate: LLAM14310 row: k column: 21  
High quality sequence stop: 433.

## FEATURES

source

Location/Qualifiers  
1. 828  
/organism="Danio rerio"  
/mol\_type="mRNA"  
/db\_xref="taxon:7955"  
/clone="IMAGE:6797422"  
/lab\_host="DH10B (T1-resistant)"  
/clone\_id="NCI CGAP ZKId1"  
/note="Organ: Kidney; Vector: PCMV-SPOrt6.1; Site 1: EcoRV; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.8 kb. Constructed by J. Wang (Research Genetics, Invitrogen Corp) from tissue donated by L. Zon (Harvard University). Note: this is a NCI CGAP library."

## ORIGIN

Query Match 3.5%; Score 103.2; DB 14; Length 828;  
Best Local Similarity 60.5%; Pred. No. 2.4e-18;  
Matches 187; Conservative 0; Mismatches 119; Indels 3; Gaps 1;  
2465 ATGCCAAACATCATTTCTAGAGTGAAGATGACACCGGAGAAATGAAAAATCG 2524  
521 ATGCTGTGATGTTTCACGGGAGAGACAGCAGCACTGACACAGTTAAACATCG 462  
2525 TGGATGTGSCATCGGCAAGACAGGCAAACTCTTTTACACCGCATATGTGCCAAG 2584  
461 TGCACTGTGGACACGAAAGCTCGTNTGAATCTCTCTCCCTGCTGCATCAGA 402  
2585 CTGACGAGATTAATGTGCAGATGTAACATATTGGACCCATCACCATTAGAAATGGCG 2644  
401 GCCACGATGTGTTGTGACAGATGTGCCAATGTTTGAAGCCATGATGAGAGTGGCG 342  
2645 GTT---TCAACCGGTAAACCCACCGCAACGAGCAACAAATATAGAAAGTCAAGCTGCTT 2701  
341 GATCCCACTCAACCAATCCAGAGAAACGACGTGACAAATAGAGTAGAGTGTG 282  
2702 GAGGATTGAATCAGTCCGTTGCCACTTAACCAATCATCAATCTTGCATTGGGTGGAA 2761  
281 GAATCTTGACAGAGACAGTAACAGATTTCCACACAGTCCCGCACAGGGTTAGAA 222  
2762 GCAATTCA 2770  
221 AGAACCGCA 213

## RESULT 14

BM039803/c 685 bp mRNA linear EST 19-OCT-2002  
LOCUS BM039803  
DEFINITION BM039803 Nori Satoh unpublished cDNA library, blood cells Ciona intestinalis cDNA clone cibd041k19 5', mRNA sequence.  
ACCESSION BM039803  
VERSION BM039803.1 GI:24140498  
KEYWORDS EST.  
ORGANISM Ciona intestinalis  
SOURCE Ciona intestinalis  
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona; Phlebobranchia; Clonidae; Ciona.  
REFERENCE 1 (bases 1 to 685)  
AUTHORS Satou, Y., Satake, M., Azumi, K., Nonaka, M., Shin-i, T., Kohara, Y. and Satoh, N.  
TITLE Expressed genes in Ciona intestinalis (2002)

JOURNAL  
COMMENT

Unpublished (2002)  
Contact: Nori Satcho  
Department of Zoology  
Kyoto University  
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan  
Tel: 81-75-753-4081  
Fax: 81-75-705-1113  
Email: satcho@ascidian.zool.kyoto-u.ac.jp.

## FEATURES

## source

Location/Qualifiers  
1..685  
/organism="Ciona intestinalis"  
/mol\_type="mRNA"  
/db\_xref="taxon:7719"  
/clone="cibd041k19"  
/tissue\_type="blood cells"  
/clone\_1ib="Nori Satcho unpublished cDNA library, blood cells"

## ORIGIN

Query Match 3.3%; Score 96; DB 13; Length 685;  
Best Local Similarity 56.5%; Pred. No. 2.8e-16;  
Matches 199; Conservative 0; Mismatches 150; Indels 3; Gaps 1;

QY 2474 ATCAATTTCTAGAGTGAAGATGACACCGGAGAAAGATTGAAAAATCGTGATGTG 2533  
DB 616 ATTTGTTCCAGTGAAGTAAATGAGCACTGACACGTTGAGAAATCGTGACAGTC 557  
QY 2534 GCACTGGCGAAAGCAGCGGAAACTCTTTTACAGCCCATATGCGCAAGCTGACGAGA 2593  
DB 556 GCTCCCTCGGTAGCCCTTCTGAAATCGGTTTGTGCGGATATGCGCAAGATACGATG 497  
QY 2594 GTATTGTCACAGTATGATCAATATTTGGACCCCATGACCATAGGAATGCGGTTTCAAC 2653  
DB 496 GTGTTTGATGAGAAATTCATATGTTGGCCCATGACATGCGGATGCGGTTTACA 437  
QY 2654 GGTAAACCCACG---GCAACGAGACCAACATATATGAAATGACACCGTCTTGAGATTGA 2710  
DB 436 GTCAATGCGCCGCTCACCAACAGATGAGCGTGAAGGAGTGTATGAGATGACTG 377  
QY 2711 ATCACTGCCGTTGCCCATTAACCATCATCATCTCTGCAATGGTGGAGCAAAATTGA 2770  
DB 376 AAGATGAGAGTGAAGATTCCTATGACAGAGCCCGCATGATTTATCAGAGAGGAG 317  
QY 2771 AATGAGACTTGGCTTGATGCGCGTGGCCATTTAACCGCTGCCGACCA 2822  
DB 316 AAAAAAGTGTGTGATTTGGCGGTGAGGACTTTGAGAGAGTTGCCAAGCA 265

## RESULT 15

BM039194/C

LOCUS BM039194 Nori Satcho unpublished cDNA library, blood cells Ciona  
DEFINITION intestinalis cDNA clone cibd039004 5', mRNA sequence.

ACCESSION BM039194

VERSION BM039194.1 GI:24139889

KEYWORDS EST

SOURCE Ciona intestinalis

ORGANISM Ciona intestinalis

Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;

Phlebobranchia; Clonidae; Ciona.

1 (bases 1 to 686)

Satcho, Y., Satake, M., Azumi, K., Nonaka, M., Shin-i, T., Kohara, Y. and

Satcho, N. Expressed genes in Ciona intestinalis (2002)

TITLE Unpublished (2002)

JOURNAL Contact: Nori Satcho

COMMENT Department of Zoology

Kyoto University

Sakyo-ku, Kyoto, Kyoto 606-8502, Japan

Tel: 81-75-753-4081

Fax: 81-75-705-1113

Email: satcho@ascidian.zool.kyoto-u.ac.jp.

FEATURES

Location/Qualifiers

## source

1..686  
/organism="Ciona intestinalis"  
/mol\_type="mRNA"  
/db\_xref="taxon:7719"  
/clone="cibd039004"  
/tissue\_type="blood cells"  
/clone\_1ib="Nori Satcho unpublished cDNA library, blood cells"

## ORIGIN

Query Match 3.3%; Score 96; DB 13; Length 686;  
Best Local Similarity 56.5%; Pred. No. 2.8e-16;  
Matches 199; Conservative 0; Mismatches 150; Indels 3; Gaps 1;

QY 2474 ATCAATTTCTAGAGTGAAGATGACACCGGAGAAAGATTGAAAAATCGTGATGTG 2533  
DB 615 ATTTGTTCCAGTGAAGTAAATGAGCACTGACACGTTGAGAAATCGTGACAGTC 556  
QY 2534 GCACTGGCGAAAGCAGCGGAAACTCTTTTACAGCCCATATGCGCAAGCTGACGAGA 2593  
DB 555 GCTCCCTCGGTAGCCCTTCTGAAATCGTTTGTGCGGATATGCGCAAGATACGATG 496  
QY 2594 GTATTGTCACAGTATGATCAATATTTGGACCCCATGACCATATGGAATGCGGTTTCAAC 2653  
DB 495 GTGTTTGATGAGAAATTCATATGTTGGCCCATGACGATGCGGATGCGGTTTACA 436  
QY 2654 GGTAAACCCACG---GCAACGAGACCAACATATATGAAATGACACCGTCTTGAGATTGA 2710  
DB 435 GTCAATGCGCCGCTCACCAACAGATGAGCGTGAAGGAGTGTATGAGATGACTG 376  
QY 2711 ATCACTGCCGTTGCCCATTAACCATCATCATCTCTGCAATGGTGGAGCAAAATTGA 2770  
DB 375 AAGATGAGAGTGAAGATTCCTATGACAGAGCCCGCATGATTTATCAGAGAGGAG 316  
QY 2771 AATGAGACTTGGCTTGATGCGCGTGGCCATTTAACCGCTGCCGACCA 2822  
DB 315 AAAAAAGTGTGTGATTTGGCGGTGAGGACTTTGAGAGAGTTGCCAAGCA 264

Search completed: July 31, 2004, 16:59:42  
Job time : 4830 secs

This Page Blank